

University of Groningen

Cohortonderzoek COOL5-18 Technisch rapport meting VO-3 in 2014

Zijsling, Djurre; Keuning, Jos; Keizer-Mittelhaeuser, Marie-Anne; Naaijer, Harm;
Timmermans, Anneke

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Zijsling, D., Keuning, J., Keizer-Mittelhaeuser, M-A., Naaijer, H., & Timmermans, A. (2017).
Cohortonderzoek COOL5-18 Technisch rapport meting VO-3 in 2014. GION onderwijs/onderzoek.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Cohortonderzoek COOL5-18

Technisch rapport meting VO-3 in 2014

Djurre Zijsling | Jos Keuning | Marie-Anne Keizer-Mittelhaëuser |
Harm Naayer | Anneke Timmermans



Cohortonderzoek COOL⁵⁻¹⁸

Technisch rapport meting VO-3 in 2014

Djurre Zijsling | Jos Keuning | Marie-Anne Keizer-Mittelhaëuser | Harm Naayer | Anneke
Timmermans

COLOFON

Het Cohortonderzoek Onderwijsloopbanen (COOL⁵⁻¹⁸) wordt in het voortgezet onderwijs uitgevoerd door GION en Cito in opdracht van de Programmaraad voor het Onderwijsonderzoek (PROO).

Bij vermelding van dit rapport in een publicatie dient gerefereerd te worden aan:

Zijsling, D., Keuning, J., M.-A. Keizer-Mittelhaeuser, Naayer, H. & Timmermans, A. C. (2017). *Cohortonderzoek COOL⁵⁻¹⁸: Technisch rapport meting VO-3 in 2014*. Groningen: GION Onderwijs/Onderzoek.

Projectmanagement

Roel Bosker, Anneke Timmermans, Jos Keuning

Werving

Harm Naayer

Logistieke ondersteuning

Service Unit, Cito

Ontwerp vragenlijsten

Hans Kuyper

Ontwerp toetsen

Ger Limpens, Karin Bügel, Rianne Voesten

Psychometrie en methodologie

Jos Keuning, Harm Naayer, Djurre Zijsling

Redactie

Djurre Zijsling, Anneke Timmermans, Jos Keuning

Eindredactie databestand

Djurre Zijsling

ISBN 978-90-367-9444-2 (eBook)

© April 2017. GION Onderwijs/Onderzoek

Niets van deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van de Directeur van het Instituut.

INHOUDSOPGAVE

1	INLEIDING	1
2	DATAVERZAMELING	3
2.1	WERVING	3
2.2	MATERIAAL EN PROCEDURE	8
2.3	REPRESENTATIVITEIT EN RESPONS	10
3	INTELLIGENTIE	19
3.1	INLEIDING	19
3.2	DE NSCCT-VERSIE VOOR VO3	19
3.3	CONSTRUCTIE INTELLIGENTIESCORES	21
3.3.1	<i>Resultaten analyses per item en per onderdeel</i>	22
3.3.2	<i>Resultaten overall analyse</i>	24
4	BEGRIJPEND LEZEN	27
4.1	MATERIAAL	27
4.2	ONDERZOEKSDSIGN	27
4.3	MEETMODEL	28
4.4	KALIBRATIE	30
4.5	MEETNAUWKEURIGHEID	32
4.6	TOETSRESULTATEN	34
4.7	AANVULLEND ONDERZOEK IN LIMBURG	35
4.8	KOPPELING MET HET CITO VOLGSYSTEEM VO	40
5	TAALVERZORGING	43
6	WISKUNDE	53
7	ENGELS	65
8	BURGERSCHAPSCOMPETENTIES	73
8.1	INLEIDING	73
8.2	SCHALEN	75
9	LEERLINGVRAGENLIJST	77
9.1	INLEIDING	77
9.2	PROCEDURES	77
9.2.1	<i>Invulinstructie</i>	77
9.2.2	<i>Ontbrekende waarden</i>	78
9.2.3	<i>Dubbele coderingen</i>	78
9.2.4	<i>Constructie schaalscores</i>	78
9.3	SYSTEMATISCHE BESCHRIJVING	79
9.3.1	<i>Welbevinden met docenten</i>	79
9.3.2	<i>Lestijd</i>	80
9.3.3	<i>Persoonlijkheid</i>	80
9.3.4	<i>Welbevinden met klasgenoten</i>	82
9.3.5	<i>Motivatie</i>	83
9.3.6	<i>Self-efficacy</i>	84

9.3.7	<i>Taakoriëntatie</i>	85
10	OUDERVVRAGENLIJST	87
10.1	INLEIDING	87
10.2	PROCEDURES	88
10.2.1	<i>Invulinstructie</i>	88
10.2.2	<i>Omgaan met onvolledige en/of inconsistente antwoorden</i>	89
10.2.3	<i>Categoriseren open antwoorden</i>	89
10.3	SYSTEMATISCHE BESCHRIJVING	90
10.3.1	<i>Achtergrondkenmerken</i>	90
10.3.2	<i>Overige variabelen</i>	99
10.3.3	<i>Constructie SOCETN (sociaal-etnische achtergrond)</i>	101
11	BESCHRIJVING DATABESTAND	105
	LITERATUUR	113
	BIJLAGEN	117
1	DIENSTROOSTER EETCAFÉ DE LOMMER	117
2	DE SLIMME TABAKSFABRIKANT	118

1 INLEIDING

Het CohortOnderzoekOnderwijsLoopbanen (COOL⁵⁻¹⁸) wordt in opdracht van de ProgrammaRaad voor het OnderwijsOnderzoek (PROO) uitgevoerd door een consortium bestaande uit Cito, GION, ITS en Kohnstamm Instituut met het CBS als extra betrokken partij. In COOL⁵⁻¹⁸ worden leerlingen gevolgd in hun schoolloopbaan vanaf groep 2 van de basisschool tot en met het eindexamen HAVO/VWO, dan wel het tweede jaar in het MBO. ITS en Kohnstamm Instituut zijn verantwoordelijk voor het veldwerkonderzoek in het primair onderwijs; Cito en GION voeren het veldwerkonderzoek uit in het voortgezet onderwijs en het middelbaar beroepsonderwijs. In de fase voorafgaand aan het veldwerkonderzoek, levert het CBS informatie over de meest recente plaats in het onderwijsstelsel van de leerlingen die eerder hebben deelgenomen. In de fase na het veldwerkonderzoek is het CBS verantwoordelijk voor het zogenaamde verRINnen van de veldwerkbestanden, waardoor het mogelijk wordt deze veldwerkbestanden binnen de beveiligde omgeving van het CBS te koppelen aan de onderwijsnummerbestanden en andere door het CBS beheerde bestanden.

In dit rapport wordt een technische beschrijving gegeven van de gegevens die als onderdeel van COOL⁵⁻¹⁸ in het schooljaar 2013/2014 zijn verzameld in de derde klas van het voortgezet onderwijs. Er zijn in de derde klas van het voortgezet onderwijs twee soorten gegevens verzameld, namelijk toetsgegevens en vragenlijstgegevens. De toetsgegevens hebben betrekking op de onderdelen intelligentie, begrijpend lezen, taalverzorging, wiskunde en Engels. De vragenlijstgegevens vallen uiteen in een leerlinggedeelte en een oudergedeelte. Ten slotte was er het onderdeel burgerschapscompetenties dat enerzijds een aantal toetsitems bevat om de kennis op dit terrein te meten, en anderzijds items waarmee de attitudes van de leerlingen in kaart zijn gebracht.

In Hoofdstuk 2 van het onderhavige rapport staat de dataverzameling in VO-3 centraal. Er wordt ingegaan op de werving van scholen, de gewenste gang van zaken bij de afname van de toetsen en vragenlijsten, en de representativiteit van de groep leerlingen die aan deze dataverzameling van COOL⁵⁻¹⁸ in het voortgezet onderwijs hebben meegedaan. In Hoofdstuk 3 wordt ingegaan op de intelligentietest (de NSCCT). In hoofdstukken 4 tot en met 7 wordt vervolgens uitgebreid ingegaan op de inhoud en psychometrische eigenschappen van respectievelijk de toetsen tekstbegrip, taalverzorging, wiskunde en Engels. Hoofdstuk 8 is gewijd aan de burgerschapscompetenties. Hoofdstukken 9 en 10 betreffen respectievelijk de leerling- en ouder vragenlijst. In Hoofdstuk 11, ten slotte, wordt een overzicht gegeven van het complete databestand dat bij het onderhavige rapport hoort.

Het databestand, welke via DANS beschikbaar is voor analyse, bevat in de eerste plaats een aantal achtergrondvariabelen. In de tweede plaats bevat het bestand (geaggregeerde) toetsscores – maar niet de itemscores – van de onderdelen intelligentie, begrijpend lezen, taalverzorging, wiskunde en Engels. In de derde plaats bevat het bestand schaa scores – maar niet de itemscores – voor het onderdeel ‘burgerschapscompetenties’. In de vierde plaats bevat het databestand zowel alle itemscores als een aantal samengestelde variabelen (voornamelijk schalen) van de vragenlijsten voor leerlingen en ouders. Hoofdstukken 2 tot en met 10 kunnen opgevat worden als de gedetailleerde verantwoording van het databestand dat in Hoofdstuk 11 beschreven wordt.

2 DATAVERZAMELING

2.1 Werving

In het schooljaar 2013/2014 heeft de derde meting van het CohortOnderzoekOnderwijsLoopbanen (COOL⁵⁻¹⁸) in de derde klassen van het voortgezet onderwijs (VO-3) plaatsgevonden. Voor deze meting was het uitgangspunt om zoveel mogelijk leerlingen in de onderzoeksgroep op te nemen die drie jaar eerder hadden deelgenomen aan de meting van COOL⁵⁻¹⁸ in groep 8 van het basisonderwijs en/of die zes jaar eerder hadden deelgenomen aan de meting van COOL⁵⁻¹⁸ in groep 5 van het basisonderwijs (de zogenaamde targetleerlingen).

De eerste meting van COOL⁵⁻¹⁸ in VO-3 heeft duidelijk gemaakt dat de oorspronkelijke doelstelling om 7000 targetleerlingen bij het onderzoek in VO-3 te betrekken (zie p. 14 onderzoeksvoorstel) te ambitieus was (Zijsling et al., 2009). Op basis hiervan is bij de tweede meting (2010/2011) de doelstelling in overleg met de PROO bijgesteld naar een streefaantal van 3100 targetleerlingen, terwijl er een absoluut minimum aantal van 1900 targetleerlingen werd vastgelegd. Met de geanticiperde deelname van de klasgenoten van deze targetleerlingen kwam het totale streefaantal op 22.500 leerlingen te liggen, en het totale minimum aantal op 15.000 leerlingen. De minimum aantallen targetleerlingen waren gebaseerd op het percentage targetleerlingen dat bij toeval op de scholen in de scholensteekproef aangetroffen zal worden (6%), en op de verwachting dat 20% van de scholen met meer dan 15 targetleerlingen aan de individuele toetsafname zou meedoen. De verwachting was dat bij het minimum aantal leerlingen rond de 110 scholen zouden moeten deelnemen, terwijl bij het nagestreefde aantal leerlingen het aantal deelnemende scholen rond de 170 zou moeten liggen. Deze bijgestelde streef- en minimumaantallen bleken in de tweede meting in 2010/2011 realistisch te zijn. In de tweede meting van COOL⁵⁻¹⁸ hebben 21.384 leerlingen deelgenomen waarvan 2.646 leerlingen als targetleerling aangemerkt konden worden. Om die reden waren het minimum aantal leerlingen (15.000) en het minimum aantal scholen (110) ook de uitgangspunten bij de start van de werving voor de derde meting.

Om de beoogde onderzoeksgroep op tijd compleet te hebben is al in mei-juni 2013 gestart met de voorbereiding van de werving. De daadwerkelijke werving heeft vervolgens plaatsgevonden via drie op elkaar afgestemde campagnes: een informerende ronde per post richting alle scholen met minimaal 10 targetleerlingen en de daarbij behorende schoolbesturen; een informerende ronde per post en per e-mail richting alle (contactpersonen op de) vestigingen die aan de eerste en/of tweede meting van COOL⁵⁻¹⁸ hadden deelgenomen; én een intensieve wervingscampagne gericht op deze twee groepen met behulp van zogenoemde ‘ambassadeurs van COOL⁵⁻¹⁸’. Vanwege gelijktijdige, concurrerende, wervingen voor grootschalige onderzoeken is in de werving voor COOL⁵⁻¹⁸ verder afstemming gezocht met de onderzoekers van Maastricht University (specifiek de onderzoeksgroep werkzaam aan de uitvoering van het Limburgse cohort Inventaar). Dit onderzoek was deels gebaseerd op COOL⁵⁻¹⁸. Er heeft afstemming plaatsgevonden waardoor een aantal Limburgse scholen konden deelnemen aan een gecombineerde COOL⁵⁻¹⁸-Inventaar afname. Deze afstemming heeft er toe geleid dat de scholen in de provincie Limburg niet door het GION zijn benaderd in de informerende wervingsronde door ambassadeurs, maar door het team van Inventaar.

Verder is de werving voor de derde meting in VO-3 gecombineerd met de werving voor de afname van de meting in vwo-6 dat eveneens in het schooljaar 2013-2014 moest plaatsvinden. Daarbij ging het voornamelijk om scholen die eerder in 2010/2011 met vo-3 klassen aan COOL⁵⁻¹⁸ hadden deelgenomen. Voor de werving voor die afname en het resultaat daarvan verwijzen we naar de bijbehorende technische rapportage (Mittelhauser-Keizer, Naaijer, Zijlsing, & Timmermans, 2015).

Vaststellen doelgroepen

In mei 2013 heeft het CBS de verdeling van de targetleerlingen (uit de eerdere afnames in PO-5 in 2007/08 en PO-8 in 2010/11) over scholen in het voortgezet onderwijs aangeleverd (op Brinnummer van de hoofdvestiging)¹. Door deze gegevens kon worden vastgesteld dat de targetleerlingen verspreid waren over 575 scholen (inclusief de provincie Limburg), en werd duidelijk om welke schoolbesturen het ging. In totaal bevatte de lijst 16.098 targetleerlingen.

Van de 575 scholen hadden er 375 tien of meer targetleerlingen (totaal 15.301). Door de koppeling van de brinnummers (Brin4) van deze scholen met de brin- en vestigingsnummers (Brin6) van de eerste en tweede rondes van COOL⁵⁻¹⁸ (inclusief de reeds afgeronde gegevensverzamelingen in Havo-5 (2010 en 2013) en vwo-6 (2011) werd duidelijk of een school eerder aan COOL⁵⁻¹⁸ had deelgenomen.

Het aantal scholen met 9 of minder targetleerlingen bedroeg 200, waarvan het grootste deel (183) tot dusver ook nooit deelgenomen had aan COOL⁵⁻¹⁸. In totaal zaten op die 200 scholen slechts 797 targetleerlingen. Onderzoeksmatig was dit derhalve de minst interessante groep. Bovendien werd verwacht dat deze scholen gezien de onbekendheid met COOL⁵⁻¹⁸ het moeilijkst te overtuigen zouden zijn om deel te nemen aan (alleen) de derde meting.

Informerende wervingsronde

De informerende wervingsronde vond vervolgens plaats via twee kanalen. Allereerst waren er 224 scholen uit de CBS lijst met 10 of meer targetleerlingen die nog nooit eerder aan COOL⁵⁻¹⁸ deelnamen. In totaal ging het op een potentieel van 7.405 targetleerlingen. Eind mei/begin juni 2013 heeft deze groep scholen een informatiepakket over COOL⁵⁻¹⁸ ontvangen. Dit informatiepakket was gericht aan de directie en de teamleiders van leerjaar 3, en bestond uit:

- een brief met toelichting over het onderzoek, de selectie van de school en een oproep tot deelname;
- een folder met daarin de belangrijkste informatie over het onderzoek;
- een voorbeeld van een schoolrapportage zoals de school die, na deelname, van het Cito retour zou ontvangen;
- een aanbevelingsbrief van de staatssecretaris van Onderwijs

De tweede groep die geïnformeerd is over de derde meting van COOL⁵⁻¹⁸ bestond uit alle contactpersonen van deelnemende vestigingen die aan de eerste en/of tweede meting hebben deelgenomen alsook de contactpersonen van vervolgmetingen in havo-5 (2010 en 2013) en vwo-6 (2011). In totaal zijn bovengenoemde informatiematerialen (digitaal en per post) aan contactpersonen van 218 vestigingen (166 scholen) toegestuurd. Op deze 166 scholen zat een potentieel van maximaal 7.933 targetleerlingen. Op 17 van deze vestigingen zaten overigens geen of minder dan 9 targetleerlingen. Deze 17 scholen zijn echter bewust in de wervingsgroep gelaten gezien een zekere mate van relatiebelang tussen scholen en

¹ De targetleerlingen op de lijst konden dus verspreid zijn over één of meerdere vestigingen van de school.

onderzoekers, en de daarbij behorende verwachte deelnamebereidheid (voor het succes van de werving in 2010/2011 zie Zijsling et al, 2012).

In totaal zijn dus 442 scholen/vestigingen geïnformeerd; 224 op schoolniveau en 218 expliciet op vestigingsniveau omdat daar al nadere contactinformatie van voorhanden was. Binnen deze 2 groepen behoren ook de scholen in Limburg. Na de informatieronde zijn deze echter niet actief benaderd door ambassadeurs of medewerkers van GION onderwijs/onderzoek. De scholen in deze provincie worden hieronder apart behandeld.

Naast de bovengenoemde groepen scholen hebben ook de schoolbesturen het betreffende pakket toegestuurd gekregen. In de daarbij behorende brief werd melding gemaakt van het feit dat één of meerdere scholen onder het bestuur wellicht nader benaderd zou worden voor het onderzoek.

Net als bij de tweede meting werd voor deze informerende wervingsronde gebruik gemaakt van speciaal ontworpen kantoormaterialen (papier, enveloppen met COOL⁵⁻¹⁸-logo). In de brieven werd nadruk gelegd op het feit dat op de school leerlingen aanwezig waren die eerder in het basisonderwijs hadden deelgenomen aan de eerste en/of tweede meting van COOL⁵⁻¹⁸, en dat zij vermoedelijk in het komende schooljaar in het derde leerjaar zouden zitten. In de brieven werd nadrukkelijk het relatief hoge aantal targets vermeld, en dat de school om die reden nader zou kunnen worden benaderd door een ambassadeur of een medewerker van het GION.

Persoonlijke werving via ambassadeurs

Vergelijkbaar met de tweede meting is gebruik gemaakt van een pool van ruim 20 personen met geruime ervaring in het onderwijsveld (oud-directeuren, docenten, teamleiders) voor de werving van het onderzoek (de zogenaamde ambassadeurs). Nadat zij eerst schriftelijk en telefonisch waren voorzien van informatie over het onderzoek is in begin september 2013 in Groningen een startbijeenkomst georganiseerd. Tijdens deze bijeenkomst werden de ambassadeurs geïnformeerd over: de opzet en het belang van het onderzoek; de wervingsbenaderingen die men zou kunnen toepassen; en met de organisatorische en financiële afwikkeling van hun inzet (naast een initiële en een onkostenvergoeding volgde er ook een financiële tegemoetkoming wanneer de werving van de school zou leiden tot daadwerkelijke deelname).

Voorafgaand aan de bijeenkomst konden de ambassadeurs op lijsten aangeven met welke scholen zij in contact stonden; in welke regio men actief was geweest of in welke regio(s) men actief wilde gaan werven; welke schoolbesturen men kende etc. Op basis van deze informatie konden ongeveer 320 van de 442 scholen/vestigingen worden toegewezen aan een ambassadeur, waarbij in ieder geval de scholen die eerder deelnamen aan COOL⁵⁻¹⁸ zijn verdeeld. Medewerkers van het GION stonden daarnaast (onder andere via eerder onderzoek) in persoonlijk contact met ongeveer 20 scholen. Met een totaal van 340 scholen is vervolgens de tweede, persoonlijke benaderingsronde van de werving gestart.

Bijna alle 340 scholen zijn in de periode van half september tot half december persoonlijk benaderd via telefoon, e-mail, en in sommige gevallen bezoek aan de school. Omdat het resultaat half december tegenviel, is er tussen half januari en eind februari een doorstart gemaakt met de werving, waarbij een extra set van 50 scholen uit de groep van 442 is benaderd en waarin scholen een minimum afnamevariant of een Cito volgsysteem variant van de afname is aangeboden (zie onder). De persoonlijke benadering heeft zich derhalve gericht op 390 van de 442 scholen/vestigingen. Na een kleine uitval hebben uiteindelijk 19 ambassadeurs een succesvolle bijdrage kunnen leveren aan de werving van scholen: een groot deel van het

wervingsresultaat (hieronder) kan worden toegeschreven aan hun enthousiaste, vasthoudende en vooral persoonlijke inzet (bezoek aan schooldirecteuren, gebruik maken van netwerkcontacten).

Wervingsproces.

Uit het wervingsproces voor de derde meting blijkt dat een goede informatievoorziening (informerende wervingsronde, website, aansprekende en nuttige materialen (schoolrapport e.d.) in combinatie met de inzet van ambassadeurs net als in de tweede meting een goede strategie is voor de weving van grootschalig onderzoek als COOL⁵⁻¹⁸. Het succes van deze wervingscampagne bleef echter achter met die van de tweede meting. Ten opzichte van de tweede meting was de deelnamebereidheid om diverse redenen lager dan verwacht. Hieronder wordt de totstandkoming van het eindresultaat beschreven vanuit drie aanmeldingstrajecten (eigen aanmelding, aanmelding na persoonlijk contact, en aanmelding via Inventaar) en twee aanpassingen aan het originele afnamedesign.

Eigen aanmelding

In september, na de informerende ronde, had slechts één school zich uit eigen beweging voor het onderzoek aangemeld. In de periode tot de kerstvakantie 2013 zijn er drie scholen geweest die, in één geval na het opvragen van meer informatie, zich zelf hebben aangemeld. Twee van die scholen hebben zich ook weer afgemeld. Uiteindelijk zijn er dus maar twee scholen die zich zonder tussenkomst van een ambassadeur of GION-medewerker hebben aangemeld voor het onderzoek. Dit staat in schril contrast met het aantal van 15 scholen dat zich voor de tweede meting van COOL⁵⁻¹⁸ uit eigen beweging aanmeldde.

Werving via persoonlijke benadering

De persoonlijke benadering heeft veel scholen opgeleverd. De ambassadeurs gebruikten hiervoor informatie dat was opgenomen in een gestandaardiseerd logboek. Hierin waren opgenomen: contactgegevens school, (eventueel) gegevens van de contactpersoon van de meest recente deelname aan een vo-3 afname van COOL⁵⁻¹⁸, verwacht aantal (target) leerlingen. De ambassadeurs vulden gedurende de wervingsactiviteiten de logboeken aan met: een omschrijving van de pogingen scholen te werven; bij een weigering, de reden(en) niet deel te nemen; en bij deelname de gemaakte afspraken met de school en de contactgegevens van de persoon die voor de afname als verdere contactpersoon zou fungeren.

In vergelijking met de werving voor de tweede meting is het eindresultaat met veel meer moeite tot stand gekomen. Naast een algemene tendens van een dalende bereidheid ‘überhaupt’ deel te nemen aan onderzoek (‘de leerlingen en docenten zijn onderzoeks- en toetsmoe’ was een veelgehoorde klacht) speelden voor veel scholen uit de tweede groep (eerder aan COOL 5-18 deelgenomen) ook de ervaringen van de eerdere deelname mee om in deze ronde af te zien van hernieuwde deelname. Overwegingen daarbij waren:

- er zijn te veel verplichte onderdelen en/of de afnamen nemen te veel onderwijstijd in beslag;
- er zijn onderdelen die we als school niet bruikbaar of interessant vinden;
- we nemen ook al vergelijkbare (Cito) toetsen af; deelname aan COOL⁵⁻¹⁸ geeft ons geen extra informatie; we willen de leerlingen niet belasten met nog meer toetsen
- de concrete opbrengsten wegen niet op tegen de omvang van de benodigde inzet voor leerlingen en docenten.

Aanbod minimumvariant afname

Om tot een voldoende respons te komen werd gedurende de loop van de werving duidelijk dat het onderzoek aan bovenstaande overwegingen tegemoet zou moeten komen. In eerste instantie is daarom

besloten scholen die twijfelden aan deelname een beperkter afnamepakket aan te bieden. Dit minimum pakket bestond voor leerlingen uit de toetsen tekstbegrip en taalverzorging Nederlands, de toets wiskunde en de afname van de leerlingvragenlijst (4 lesuren). De NSCCT en de toets Engels/vragenlijst burgerschap werden daarmee niet meer verplicht. In totaal hebben acht geworven scholen gekozen voor deze minimum variant van de afname. Deze scholen ontvingen weliswaar allen de oudervragenlijsten, maar deze is vervolgens niet op alle zeven scholen afgenomen.

Aanbod delen toetsscores uit Cito volgsysteem

Veel scholen in de wervingspool namen, naar eigen zeggen, ook de toetsen van het Cito volgsysteem voortgezet onderwijs af. Dit bleek een groot effect te hebben op de bereidheid van scholen om deel te nemen aan een onderzoek als COOL⁵⁻¹⁸. Zicht op de prestaties van leerlingen werd immers al via het eigen leerlingvolgsysteem verkregen en de COOL⁵⁻¹⁸-toetsen zouden daar geen extra informatie voor opleveren. Om deze reden is bij de derde meting, bij wijze van proef, toegestaan dat de toetsen van het Cito volgsysteem voortgezet onderwijs als alternatief zouden worden gebruikt voor de COOL⁵⁻¹⁸-toetsen. Deelname aan COOL⁵⁻¹⁸ hield dan in: het toestemming geven om de scores op de toetsen van de VO-3 leerlingen uit het Cito Volgsysteem VO te gebruiken (Nederlands, wiskunde, Engels); en afname van de ouder- en leerling vragenlijsten. De onderdelen NSCCT en de vragenlijst burgerschap vervielen in dit scenario ook tenzij de school aangaf deze wel te willen afnemen. In totaal hebben negen scholen voor deze variant gekozen.

Werving Limburg

Aangezien in de provincie Limburg een op COOL⁵⁻¹⁸ geënt onderzoek plaats vond, was het niet opportuun om daar zelfstandig te gaan werven. In goed overleg met Inventaar zijn door de veldwerk coördinator van dit instituut uiteindelijk 10 scholen met relatief veel target leerlingen bereid gevonden deel te nemen aan de gecombineerde afname van het Inventaar/ COOL⁵⁻¹⁸ instrumentarium.

Eindresultaat

De persoonlijke werving via ambassadeurs en GION medewerkers heeft tussen half september en half december 2013 86 scholen opgeleverd. Voorts zijn tussen half januari en eind februari 2014 nog eens 23 scholen geworven waarbij de nadruk lag op het aanbod van de minimum variant of de Cito-volgsysteem variant van de afname. Het resultaat van de persoonlijke benadering is dus 109 scholen. Uitgaande van 390 persoonlijk benaderde scholen is het netto resultaat van de persoonlijke benadering derhalve 27,9%.

Tezamen met de scholen die zich zelf aanmelden (2) en de 10 Limburgse scholen die werden aangemeld door Inventaar voor de collectieve afname van het gezamenlijke Inventaar/ COOL⁵⁻¹⁸ instrumentarium bestond het eindresultaat van de werving uit 121 scholen. 79 daarvan zouden collectief (met alle onderdelen) deelnemen, 8 met een minimum variant en 9 scholen met de Cito volgsysteem variant. Daarnaast zouden 15 van de 109 scholen op individuele basis gaan deelnemen: op deze scholen zouden alleen de targetleerlingen aan de afname (van alle onderdelen) meedoen.

Leerlingegevens, uitval en 'netto'verzendlst

Tussen oktober 2013 en februari 2014 zijn bij de 121 scholen die hadden toegezegd mee te zullen werken gegevens van de leerlingen opgevraagd. Deze gegevens waren nodig om op naam gestelde antwoordbladen voor de toetsen en vragenlijsten te kunnen maken. Bovendien kon op basis van de leerlingegegevens een meer precieze inschatting gemaakt worden van het totale aantal deelnemende leerlingen. Aan de scholen die zouden deelnemen, is gevraagd om de volgende gegevens digitaal naar het GION te sturen:

1. Naam van de leerling (voornaam en achternaam)
2. Klas (de aanduiding, zoals die op de school gebruikt wordt)
3. Geboortedatum
4. Geslacht
5. Type onderwijs dat de leerling volgt (bbl, kbl, tl, gl, gtl, havo, havo/vwo, vwo)

Tussen oktober 2013 en februari 2014 zijn alsnog 11 van de 121 scholen (9,1%) uitgevallen (negen collectief, één individueel deelnemende, en één school die de Cito volgsysteem variant zou). De voornaamste reden daarvoor was dat ondanks dat een contactpersoon deelname had toegezegd er (later) onvoldoende capaciteit en draagvlak binnen de school was om het onderzoek daadwerkelijk uit te voeren en te coördineren. Eind februari bedroeg de definitieve ‘verzendlst’ 110 scholen, waarvan er 81 collectief zouden deelnemen, 8 de minimum variant zouden afnemen, 8 de Cito-volgsysteem variant, 10 de Inventaar/COOL⁵⁻¹⁸ variant en 14 ‘individueel’. Gezamenlijk zaten op deze scholen 19.817 leerlingen en voor dit aantal leerlingen zijn dan ook de materialen voor de zending aangemaakt.

Vergelijkbaar met de gegevensverzameling van de eerste meting (2007/2008), maar afwijkend van de tweede meting (2010/2011) zijn in deze derde meting gekozen alle toetsmaterialen en vragenlijsten in één keer naar de school gezonden. De scholen konden hierdoor over alle materialen beschikken en de afname naar eigen inzicht verspreid over de tweede helft van het schooljaar organiseren (zie paragraaf 2.2).² De zendingen hebben wel op twee momenten plaatsgevonden: het grootste deel ontving alle materialen in begin februari, de rest van de scholen (die zich later hadden aangemeld) in begin maart.

2.2 Materiaal en procedure

In de derde afname van COOL⁵⁻¹⁸ in het derde leerjaar van het voortgezet onderwijs is net als in de voorgaande metingen een aantal vaardigheden getoetst, namelijk de cognitieve capaciteit (intelligentie) en de schoolse vaardigheden begrijpend lezen, taalverzorging, wiskunde en Engels.³ Daarnaast is bij een deel van de leerlingen een vragenlijst met betrekking tot burgerschapscompetenties afgenomen. De opzet was dat leerlingen die aan het onderdeel burgerschapscompetenties deelnamen, de toets Engels niet zouden maken. Scholen mochten in principe zelf aangeven of ze leerlingen de toets Engels of het onderdeel burgerschapscompetenties voor wilden leggen. Via een e-mail ronde konden zij hun keuze vastleggen. Coördinatoren die niet reageerden werden random aan één van de onderdelen toegewezen, maar in veel gevallen hadden zij zelf al geen voorkeur. Zeventien scholen wilden zowel Engels als burgerschapscompetenties afnemen. Alle toetsen hadden het meerkeuze formaat met doorgaans vier antwoordmogelijkheden per item. De leerlingen moesten hun antwoorden op aparte antwoordbladen aanstrepen. In de hoofdstukken 3 tot en met 8 wordt uitgebreid ingegaan op de inhoud en (psychometrische) eigenschappen van de afgenomen toetsen.

² De gespreide zending die in de tweede meting werd toegepast leverde scholen soms logistieke problemen op (bijvoorbeeld omtrent het bewaren van reeds afgenomen onderdelen), als ook problemen rondom het inplannen van de afname (er was bijvoorbeeld na de derde zending geen lesuren meer in te roosteren voor alle onderdelen uit die zending).

³In de eerste meting (2008) bestond het onderdeel taalverzorging uit twee afzonderlijke deeltoltsen, namelijk woordenschat en werkwoordspelling.

Naast de genoemde toetsen zijn ook twee vragenlijsten afgenomen, namelijk een leerlingvragenlijst en een oudervragenlijst. De leerlingvragenlijst is vooral afgenomen om de sociaal-emotionele ontwikkeling van de leerlingen in kaart te brengen. De oudervragenlijst was vooral bedoeld om enkele achtergrondkenmerken vast te stellen. Bijna alle vragen hadden gesloten antwoordmogelijkheden. In hoofdstukken 9 en 10 wordt uitgebreid ingegaan op de inhoud van respectievelijk de leerling- en oudervragenlijst.

In Tabel 2.1 zijn alle onderdelen opgesomd. In de loop van de werving (zie boven) werd duidelijk dat veel scholen niet alle onderdelen wilden afnemen en/of dat zij reeds aan de toetsen van het Cito volgsysteem deelnamen. Om die reden zijn derhalve twee varianten aan scholen aangeboden, de minimum variant (onderdelen a,b,d,e en g) en de Cito-volgsysteem variant (onderdelen a en g eventueel aangevuld met onderdeel c).

Tabel 2.1

COOL⁵⁻¹⁸ – Onderdelen, verzending en afname VO-3

Lesuur*	Onderdeel
	- antwoordbladen voor alle toetsen
	- handleiding per deelnemende klas
	a. oudervragenlijst + toelichtende brief
1	b. toets taalverzorging Nederlands
2	c. toets Engels en/of burgerschap **
3	d. toets wiskunde
4	e. toets tekstbegrip Nederlands
5	f. intelligentietest
6	g. leerlingvragenlijst

* Volgorde van de onderdelen door de school zelf te bepalen. **Afhankelijk van de keuze school

De contactpersoon bij aanmelding van de school werd aangesteld als coördinator van de uitvoeringen was degene die het onderzoeksmateriaal ontving, binnen de school verspreidde, en uiteindelijk terugstuurde naar Cito. De coördinator en de afnamebegeleider(s) (docenten) waren samen verantwoordelijk voor de daadwerkelijke toets afnemen binnen de klas. Om deze toetsafnamen goed te laten verlopen is zowel voor de coördinator als voor de daadwerkelijke afnameleider per deelnemende klas een uitgebreide handleiding naar de scholen gestuurd. Zij zorgden er voor dat de toetsboekjes uitgereikt werden, dat het juiste antwoordblad bij de juiste leerling terecht kwam, dat de antwoordbladen weer ingenomen werden, en dat er tijdens de toetsafname niet afgekeken werd.

Voor de toetsen was de volgende globale instructie opgenomen in de handleiding:

Het is belangrijk dat de afnamecondities gelijk zijn aan de afnamecondities bij een proefwerk. Een surveillerende docent moet ervoor zorgen dat de leerlingen ongestoord en individueel te werk kunnen gaan. Het maken van elke toets neemt maximaal één lesuur van 50 minuten in beslag. Elke toets heeft drie versies die qua moeilijkheid zijn afgestemd op het vaardigheidsniveau van de leerlingen in een bepaald klastype. Versie 1 is steeds bedoeld voor BBL-leerlingen, versie 2 voor KBL- en GL/TL-leerlingen, en versie 3 voor HAVO- en VWO-leerlingen. In principe krijgen alle leerlingen in een klas dezelfde versie. Alleen in geval er bij u op school gemengde derde klassen zijn (bijvoorbeeld BBL/KBL), krijgen niet alle leerlingen in een klas dezelfde versie. Om inzichtelijk te houden welke leerling welke versie moeten krijgen, zitten er nooit verschillende versies in één en dezelfde

klassendoos. Bij een gemengde klas zal er dus sowieso sprake zijn van meerdere dozen, namelijk minimaal 1 doos voor elke versie.

De specifieke instructies voor de afname per (toets)onderdeel waren in de aparte hoofdstukken opgenomen.

Het materiaal was per onderdeel per klas en per onderwijsniveau (d.w.z., BBL, KBL/GL/TL en HAVO/VWO) geordend. Een verzendoos voor een onderdeel bevatte voor elke leerling in de klas het benodigde toetsboekjes en/of de leerlingvragenlijst. De antwoordbladen voor alle vier de toetsen (taalverzorging, wiskunde, Engels, begrijpend lezen) werden meegestuurd. De oudervragenlijsten waren apart, per klas verpakt.

De antwoordbladen en vragenlijsten waren op naam gemaakt. Dat wil zeggen dat op elk antwoordblad de naam van de leerling en een bijbehorend identificatienummer stonden voorgedrukt. Als een mentor of afnamebegeleider materiaal tekort kwam, kon er daarom niet zomaar gekopieerd worden. Dan zou één en dezelfde leerling immers twee toetsresultaten of vragenlijsten toegewezen krijgen. De verzendoos bevatte daarom ook een aantal reserve antwoordbladen en vragenlijsten. Op het reservemateriaal was alleen een unieke barcode aangebracht; de gegevens van de leerling moesten handmatig ingevuld worden. Op deze wijze zijn nog 41 leerlingen toegevoegd. De unieke codering van de antwoordbladen en vragenlijsten maakte het mogelijk om alle antwoorden van de leerlingen automatisch in te lezen. Het optisch inlezen van de antwoordbladen en vragenlijsten heeft plaatsgevonden bij Cito.

De dataverzameling in VO-3 heeft plaatsgevonden vanaf begin februari tot uiterlijk eind juni 2014. De afname van de toetsen en vragenlijsten kon zeer flexibel plaatsvinden. Scholen mochten zelf de dagen kiezen waarop de toetsen en de vragenlijst zouden worden afgenomen en ondanks de verzending in drie delen was er geen volgorde aangegeven waarin de onderdelen moesten worden afgenomen. Voor de collectieve afname waren in totaal 6 lesuren nodig; maximaal 1 lesuur per toets of vragenlijst. Wanneer de school had gekozen voor zowel burgerschap als Engels betekende dat één extra afname-uur. De oudervragenlijst werd buiten de school afgenomen. De leerlingen kregen deze vragenlijst mee naar huis en moesten deze na invulling door een van hun (of beide) ouders/verzorgers in een afsluitbare retourenvelop weer inleveren. Ook bestond voor de ouders de mogelijkheid om de oudervragenlijst retour te sturen naar het antwoordnummer van het GION. Scholen werden in de handleiding verzocht om het afgenomen onderdeel en in dezelfde dozen zo snel mogelijk aan Cito te retourneren. Nadat alle scholen de materialen teruggestuurd hadden en de gegevens waren verwerkt, heeft er een terugrapportage naar de scholen plaatsgevonden. Van alle toetsen werd de toetsscore en de (relatieve) referentieniveaus teruggekoppeld. Over de vragenlijsten en het onderdeel burgerschapscompetenties is niet terug gerapporteerd.

2.3 Representativiteit en respons

In totaal staan er van 16.297 leerlingen gegevens in het databestand. De leerlingen zaten op 107 verschillende scholen. Er zijn 197 leerlingen die van hun ouders niet aan het onderzoek mee mochten doen. De ouders van deze leerlingen hebben dus gebruikgemaakt van de mogelijkheid die in de ouderbrief geboden werd om via het sturen van een e-mail naar cool@cito.nl kenbaar te maken dat we de resultaten van hun kind niet mochten gebruiken. Om organisatorische redenen hebben deze leerlingen de toetsen en vragenlijsten vaak wel gemaakt. De resultaten van deze leerlingen (de zogenaamde ‘weigeraars’) zijn

echter buiten het databestand gelaten en komen niet voor in de beschrijvingen die in dit rapport gegeven worden.

Een leerling is tot de uiteindelijke onderzoeksgroep gerekend als er ten minste gegevens beschikbaar waren van één onderdeel. Meer dan 27 procent van de leerlingen (4498 in totaal) heeft deelgenomen aan alle zeven onderdelen (of acht als scholen zowel deelnamen aan Engels als burgerschapscompetenties). Van de overige leerlingen hebben er 2822 (17.3%) deelgenomen aan zes onderdelen, 2477 (15.2%) aan vijf onderdelen, 1515 (9.3%) aan vier onderdelen, 3194 (19.6%) aan drie onderdelen, 709 (4.4%) aan twee onderdelen en 1082 (6.6%) aan één onderdeel. In het databestand geeft de variabele **NTOETS** weer aan hoeveel onderdelen een leerling heeft deelgenomen. Van 3124 leerlingen in de oorspronkelijke groep van 19.618 leerlingen (zie paragraaf 2.1) zijn van geen enkel onderdeel gegevens beschikbaar. Dit kan het gevolg zijn van (a) een weigering tot deelname door de ouders/verzorgers van de leerling (b) het afhaken van een school, en (c) afwezigheid van een leerling tijdens de toetsafname door ziekte of andere omstandigheden.

Als een leerling deelgenomen heeft aan een onderdeel betekent dat niet automatisch dat deze leerling ook alle items van dat onderdeel beantwoord heeft. In het databestand geven de variabelen **MIS_BGL**, **MIS_TVZ**, **MIS_WIS** en **MIS_ENG** per leerling aan op hoeveel procent van de items een antwoord ontbreekt. Bij de berekening van het percentage ontbrekende antwoorden is uitgegaan van het aantal items in een toets. Er is geen rekening gehouden met de weging die naar aanleiding van de analyses met het *One-Parameter Logistic Model* (zie paragraaf 4.3) is toegepast bij de scoring. Tabel 2.2 laat zien dat verreweg de meeste leerlingen alle items beantwoord hebben. Gemiddeld gezien is bij meer dan 92 procent van de leerlingen het percentage ontbrekende antwoorden gelijk aan 0. Het percentage leerlingen dat gemiddeld gezien meer dan 5 procent van de items niet beantwoord heeft, is bovendien erg klein (3.8%). In sommige situaties kan het zinvol zijn om de leerlingen met relatief veel ontbrekende antwoorden buiten beschouwing te laten. Bij veel ontbrekende antwoorden kan er immers aan getwijfeld worden of een leerling wel gemotiveerd heeft deelgenomen. Bij de berekening van de scores zijn de items met een ontbrekend antwoord fout gerekend. Tabel 2.2 bevat geen informatie over het aantal ontbrekende antwoorden bij het burgerschapsinstrumentarium en de leerling- en oudervragenlijst. De reden hiervoor is dat dit aantal per vraag varieert en er bij het maken van de variabelen schaalspecifieke regels met betrekking tot ontbrekende waarden zijn gehanteerd. Zie hiervoor de informatie in Hoofdstuk 8, 9 en 10. Ook bij de intelligentietest is op een enigszins afwijkende manier omgegaan met ontbrekende antwoorden, omdat de afname bij deze test een licht ‘speeded’ karakter had. Leerlingen hebben alleen een score toegekend gekregen indien ten minste de helft van de items gemaakt was.

Tabel 2.2

Percentage missing per onderdeel

Onderdeel	Percentage ontbrekende antwoorden							Totaal
	0	1-5	6-10	11-20	21-50	51-99	100	
Begrijpend lezen	12811	532	41	114	672	69	12	14251
Wiskunde	12663	493	117	123	191	498	0	14085
Engels ¹	8618	214	14	19	37	16	0	8918
Taalverzorging	8203	504	30	6	20	63	0	8826

Noot1: Engels is niet aan alle leerlingen in de onderzoeksgroep voorgelegd.

Zoals eerder aangegeven is geprobeerd om zoveel mogelijk leerlingen in de onderzoeksgroep op te nemen die hadden deelgenomen aan een van de eerdere afnames van COOL⁵⁻¹⁸: de afname in PO-5 in 2007/08 en/of de afname in PO-8 in 2010/11. Een eerste analyse liet zien dat 2278 van de 16.297 leerlingen ook in eerdere databestanden voor het basisonderwijs voorkomen. Van deze groep hebben 925 leerlingen (40%) het onderdeel intelligentie gemaakt, 1908 leerlingen (83%) het onderdeel begrijpend lezen, 1222 leerlingen (53.0%) het onderdeel taalverzorging, 1924 leerlingen (84%) het onderdeel wiskunde en 1190 leerlingen (52%) het onderdeel Engels. De leerling- en ouder vragenlijst zijn door respectievelijk 2075 leerlingen (91%) en 1108 ouders van leerlingen (48%) ingevuld. Aan het onderdeel burgerschapscompetenties hebben ‘slechts’ 702 leerlingen (30%) deelgenomen. Als we kijken naar het aantal onderdelen dat elke ‘target’-leerling heeft gemaakt, zien we het volgende: alle onderdelen 29% (668 leerlingen), 6 onderdelen 13% (311 leerlingen), 5 onderdelen 15% (354 leerlingen), 4 onderdelen 10% (249 leerlingen), 3 onderdelen 17% (398 leerlingen), 2 onderdelen 4% (110 leerlingen) en 1 onderdeel 8% (188 leerlingen). Het is belangrijk op te merken dat de informatie met betrekking tot de targetleerlingen nog voorlopig is. Het onderwijsnummerbestand van het CBS moet namelijk nog gekoppeld worden aan het COOL⁵⁻¹⁸ databestand voor VO-3.

Om de representativiteit van de onderzoeksgroep te kunnen beoordelen is de onderzoeksgroep geanalyseerd in het perspectief van een aantal relevante achtergrondvariabelen. De representativiteit is geëvalueerd in relatie tot de volgende variabelen: (1) onderwijstype, (2) regionale spreiding, (3) mate van verstedelijking, (4) sekse en (5) leeftijd. Voor het bepalen van de populatiegegevens is gebruikgemaakt van gegevens van DUO. Concreet is uitgegaan van het bestand met het aantal leerlingen per vestiging uitgesplitst naar onderwijstype en leerjaar in schooljaar 2013/2014. Omdat niet alle leerlingen in de onderzoeksgroep hebben deelgenomen aan alle onderdelen zijn de analyses niet alleen voor de totale onderzoeksgroep verricht (d.w.z., de leerlingen die aan ten minste één onderdeel hebben deelgenomen), maar ook voor elk van de onderdelen afzonderlijk.

Representativiteit naar onderwijstype

Eerst is gekeken naar de vertegenwoordiging van de verschillende onderwijstypen in de onderzoeksgroep. Er zijn zes onderwijstypen onderscheiden: (1) basisberoepsgerichte leerweg – BBL, (2) kaderberoepsgerichte leerweg – KBL, (3) gemengde en theoretische leerweg – GL/TL, (4) HAVO – H, (5) VWO – V en (6) HAVO/VWO – H/V. De verdeling van leerlingen in de populatie en onze onderzoeksgroep naar onderwijstype staat in Tabel 2.3.

We zien dat de verschillende leerwegen van het VMBO ondervertegenwoordigd zijn in de onderzoeksgroep en dat de onderwijstypen HAVO en VWO oververtegenwoordigd zijn. De afwijking tussen de aantallen leerlingen in onze onderzoeksgroep en de populatie zijn significant: $\chi^2 = 2033.382$, $df = 5$, $p < 0.001$, $\Phi_{\text{Cramer}} = 0.354$. Dit betekent dat het in analyses waarin de totale onderzoeksgroep geanalyseerd wordt noodzakelijk is om terug te wegen voor de variabele *onderwijstype*. In het databestand geeft de variabele **OW_TYPE** aan welk type onderwijs een leerling volgde ten tijde van de dataverzameling.

Tabel 2.3*Aantal en percentage leerlingen in de populatie en de onderzoeksgroep naar onderwijstype*

Onderdeel	Onderwijstype						Totaal (onbekend)
	BBL	KBL	GL/TL	H	H/V	V	
Onderzoeksgroep	1091 (6.7)	1487 (9.1)	3707 (22.8)	5088 (31.2)	221 (1.4)	4678 (28.7)	16297 (25)
Intelligentie	518 (7.3)	713 (10.0)	1880 (26.4)	2060 (28.9)	117 (1.6)	1825 (25.6)	7116 (3)
Begrijpend lezen	783 (5.5)	1118 (7.8)	3228 (22.7)	4705 (33.0)	214 (1.5)	4199 (29.5)	14251 (4)
Taalverzorging	543 (6.2)	811 (9.2)	1976 (22.4)	2748 (31.1)	120 (1.4)	2626 (29.8)	8826 (2)
Wiskunde	779 (5.5)	1055 (7.5)	3247 (23.1)	4709 (33.4)	215 (1.5)	4073 (28.9)	14085 (7)
Engels	338 (3.8)	573 (6.4)	1967 (22.1)	3168 (35.5)	123 (1.4)	2746 (30.8)	8918 (3)
Burgerschap	250 (5.7)	298 (6.7)	1252 (28.3)	1357 (30.7)	106 (2.4)	1153 (26.1)	4418 (2)
Leerlingvragenlijst	957 (6.6)	1318 (9.2)	3347 (23.3)	4517 (31.4)	212 (1.5)	4037 (28.0)	14394 (6)
Oudervragenlijst	482 (6.0)	672 (8.4)	1815 (22.7)	2497 (31.3)	128 (1.6)	2383 (29.8)	7989 (12)
Populatie	22260 (11.0)	30067 (14.9)	56484 (28.0)	43913 (21.8)	6152 (3.0)	42978 (21.3)	201863

Representativiteit naar regionale spreiding

In een vervolgstap is gekeken naar de spreiding van scholen en leerlingen over Nederland. Er is onderscheid gemaakt in vier regio's. Regio *Noord* omvat de provincies Groningen, Friesland en Drenthe; regio *Oost* de provincies Overijssel, Gelderland en Flevoland; regio *West* de provincies Utrecht, Noord-Holland, Zuid-Holland en Zeeland en regio *Zuid* de provincies Noord-Brabant en Limburg. De verdeling van scholen en leerlingen in de populatie en onze onderzoeksgroep naar regio staat in Tabel 2.4a en Tabel 2.4b. Merk op dat er in de analyse vanuit is gegaan dat de regio voor de scholen en de leerlingen identiek is. Dit hoeft vanzelfsprekend niet per se zo te zijn. Als een leerling ver van de school afwoont, kan een leerling in een andere regio wonen dan naar school gaan.

We zien dat de verdeling van scholen en leerlingen over de vier regio's in de onderzoeksgroep tamelijk sterk afwijkt van de verdeling in de populatie. Er hebben ten opzichte van de populatie tamelijk veel scholen en leerlingen uit het zuiden meegedaan aan COOL⁵⁻¹⁸ en tamelijk weinig scholen en leerlingen uit het westen. Als we de totale onderzoeksgroep afzetten tegen de populatie zien we dat er zowel op schoolniveau als leerlingniveau sprake is van een significant verschil tussen onze onderzoeksgroep en de aantallen scholen en leerlingen die we op grond van de gegevens van DUO in elke regio mogen verwachten: (schoolniveau) $\chi^2 = 21.094$, $df = 3$, $p < 0.001$, $\Phi_{\text{Cramer}} = 0.476$; (leerlingniveau) $\chi^2 = 4583.407$, $df = 3$, $p < 0.001$, $\Phi_{\text{Cramer}} = 0.530$. In bepaalde analyses kan het dus wenselijk zijn om terug te wegen voor de variabele *regio*. Een alternatief kan zijn om de variabele *regio* expliciet in een model op te nemen. In het databestand geeft de variabele **SCH_REG** aan in welke provincie de school stond. De hierboven gehanteerde indeling in regio's is op basis van deze variabele eenvoudig te maken. Ook kan op basis van deze variabele gemakkelijk de representativiteit naar regio binnen een onderwijstype onderzocht worden.

Tabel 2.4a*Aantal en percentage scholen in de populatie en de onderzoeksgroep naar regio*

Onderdeel	Regio				Totaal
	Noord	Oost	West	Zuid	
Onderzoeksgroep	8 (8.6)	28 (30.1)	27 (29.0)	30 (32.3)	93
Intelligentie	5 (10.4)	16 (33.3)	18 (37.5)	9 (18.8)	48
Begrijpend lezen	8 (9.1)	26 (29.5)	26 (29.5)	28 (31.8)	88
Taalverzorging	8 (12.1)	21 (31.8)	23 (34.8)	14 (21.2)	66
Wiskunde	8 (9.3)	25 (29.1)	25 (29.1)	28 (32.6)	86
Engels	8 (13.8)	14 (24.1)	19 (32.8)	17 (29.3)	58
Burgerschap	3 (9.7)	12 (38.7)	9 (29.0)	7 (22.6)	31
Leerlingvragenlijst	8 (9.1)	27 (30.7)	24 (27.3)	29 (33.0)	88
Oudervragenlijst	7 (9.0)	27 (34.6)	24 (30.8)	20 (25.6)	78
Populatie	152 (13.0)	247 (21.1)	556 (47.4)	217 (18.5)	1172

Tabel 2.4b*Aantal en percentage leerlingen in de populatie en de onderzoeksgroep naar regio*

Onderdeel	Regio				Totaal
	Noord	Oost	West	Zuid	
Onderzoeksgroep	1533 (9.4)	3705 (22.7)	4359 (26.7)	6700 (41.1)	16297
Intelligentie	602 (8.5)	2211 (31.1)	2564 (36.0)	1739 (24.4)	7116
Begrijpend lezen	1298 (9.1)	3326 (23.3)	4034 (28.3)	5593 (39.2)	14251
Taalverzorging	902 (10.2)	2487 (28.2)	3077 (34.9)	2360 (26.7)	8826
Wiskunde	1204 (8.5)	3294 (23.4)	3900 (27.7)	5687 (40.4)	14085
Engels	1296 (14.5)	1791 (20.1)	2890 (32.4)	2941 (33.0)	8918
Burgerschap	366 (8.3)	1647 (37.3)	1314 (29.7)	1091 (24.7)	4418
Leerlingvragenlijst	1288 (8.9)	3282 (22.8)	3773 (26.2)	6051 (42.0)	14394
Oudervragenlijst	753 (9.4)	2260 (28.3)	2364 (29.6)	2612 (32.7)	7989
Populatie	21716 (10.7)	45076 (22.3)	93358 (46.1)	42167 (20.8)	202317

representativiteit naar mate van verstedelijking

Vervolgens is de representativiteit van de onderzoeksgroep geanalyseerd in het licht van de achtergrondvariabele *mate van verstedelijking*. Er zijn vijf groepen onderscheiden: scholen of leerlingen in (1) niet verstedelijkt gebied, (2) weinig verstedelijkt gebied, (3) matig verstedelijkt gebied, (4) sterk verstedelijkt gebied, en (5) zeer sterk verstedelijkt gebied. De verdeling van scholen en leerlingen in de populatie en onze onderzoeksgroep naar mate van verstedelijking staat in Tabel 2.5a en Tabel 2.5b.

We zien dat de verdeling van scholen en leerlingen in onze onderzoeksgroep naar mate van verstedelijking redelijk lijkt op de verdeling in de populatie. Alleen de zeer sterk verstedelijkte gebieden zijn wat ondervertegenwoordigd. De afwijking tussen de aantallen scholen in onze onderzoeksgroep en de aantallen scholen die verwacht mogen worden op grond van de gegevens van DUO zijn niet significant: $\chi^2 = 8.384$, $df = 4$, $p = 0.078$, $\Phi_{\text{Cramer}} = 0.300$. Effectgrootte Φ_{Cramer} is echter wel behoorlijk groot. Op leerlingniveau is er sprake van een significant verschil: $\chi^2 = 1528.163$, $df = 4$, $p < 0.001$, $\Phi_{\text{Cramer}} = 0.306$. Effectgrootte Φ_{Cramer} is vergelijkbaar met die op schoolniveau en wederom als behoorlijk groot aan te merken. Dit betekent dat het in bepaalde analyses wenselijk kan zijn om terug te wege voor de variabele *mate van verstedelijking*. In het databestand geeft de variabele **SCH_STED** aan in hoeverre de school in verstedelijkt gebied stond. Merk op dat er in de analyses vanuit is gegaan dat de stedelijke graad voor de scholen en de leerlingen identiek is. Dit hoeft vanzelfsprekend niet per se zo te zijn. Als een leerling ver van de school afwoont, kan de stedelijke graad van de woonomgeving van de leerling afwijken van de stedelijke graad van de schoolomgeving.

Tabel 2.5a

Aantal en percentage scholen in de populatie en de onderzoeksgroep naar mate van verstedelijking

Onderdeel	Mate van verstedelijking					Totaal
	Niet	Weinig	Matig	Sterk	Zeer sterk	
Onderzoeksgroep	6 (6.5)	22 (23.7)	19 (20.4)	36 (38.7)	10 (10.8)	93
Intelligentie	4 (8.3)	12 (25.0)	11 (22.9)	16 (33.3)	5 (10.4)	48
Begrijpend lezen	6 (6.8)	21 (23.9)	19 (21.6)	33 (37.5)	9 (10.2)	88
Taalverzorging	5 (7.6)	14 (21.2)	13 (19.7)	25 (37.9)	9 (13.6)	66
Wiskunde	6 (7.0)	21 (24.4)	18 (20.9)	32 (37.2)	9 (10.5)	86
Engels	4 (6.9)	12 (20.7)	14 (24.1)	19 (32.8)	9 (15.5)	58
Burgerschap	2 (6.5)	7 (22.6)	9 (29.0)	11 (35.5)	2 (6.5)	31
Leerlingvragenlijst	6 (6.8)	21 (23.9)	19 (21.6)	35 (39.8)	7 (8.0)	88
Oudervragenlijst	4 (5.1)	18 (23.1)	17 (21.8)	31 (39.7)	8 (10.3)	78
Populatie	81 (6.9)	214 (18.3)	238 (20.3)	376 (32.1)	263 (22.4)	1172

Tabel 2.5b*Aantal en percentage leerlingen in de populatie en de onderzoeksgroep naar mate van verstedelijking*

Onderdeel	Mate van verstedelijking					Totaal
	Niet	Weinig	Matig	Sterk	Zeer sterk	
Onderzoeksgroep	786 (4.8)	4378 (26.9)	3920 (24.1)	5786 (35.5)	1427 (8.8)	16297
Intelligentie	285 (4.0)	1911 (26.9)	1655 (23.3)	2833 (39.8)	432 (6.1)	7116
Begrijpend lezen	709 (5.0)	3738 (26.2)	3527 (24.7)	4956 (34.8)	1321 (9.3)	14251
Taalverzorging	527 (6.0)	2167 (24.6)	2230 (25.3)	2953 (33.5)	949 (10.8)	8826
Wiskunde	686 (4.9)	3686 (26.2)	3483 (24.7)	4960 (35.2)	1270 (9.0)	14085
Engels	494 (5.5)	1881 (21.1)	2247 (25.2)	3176 (35.6)	1120 (12.6)	8918
Burgerschap	88 (2.0)	766 (17.3)	1340 (30.3)	1894 (42.9)	330 (7.5)	4418
Leerlingvragenlijst	758 (5.3)	3954 (27.5)	3519 (24.4)	5156 (35.8)	1007 (7.0)	14394
Oudervragenlijst	225 (2.8)	2060 (25.8)	2115 (26.5)	2847 (35.6)	742 (9.3)	7989
Populatie	9344 (4.6)	37515 (18.5)	49486 (24.5)	67641 (33.4)	38331 (18.9)	202317

Representativiteit naar sekse

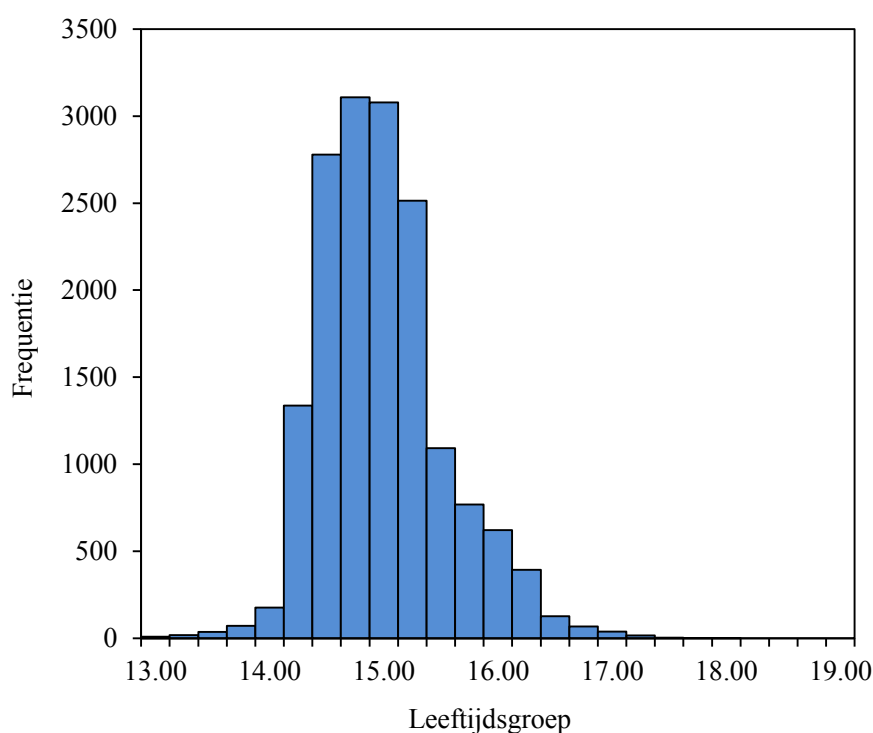
Na de beoordeling van de representativiteit van de onderzoeksgroep in het licht van de variabelen *onderwijstype*, *regio* en *mate van verstedelijking* is gekeken naar de verdeling van jongens en meisjes. De verdeling van leerlingen in de populatie en onze onderzoeksgroep naar sekse staat in Tabel 2.6. Volgens de gegevens van DUO zaten er in schooljaar 2013/2014 iets meer jongens dan meisjes in leerjaar 3 van het voortgezet onderwijs, namelijk 50.6 procent jongens tegenover 49.4 procent meisjes. Deze verhouding komt overeen met onze onderzoeksgroep. Als we de totale COOL⁵⁻¹⁸-onderzoeksgroep afzetten tegen de populatie zien we dat de aantallen jongens en meisjes in onze onderzoeksgroep niet significant afwijken van de aantallen die op grond van de gegevens van DUO verwacht mogen worden: $\chi^2 < 0.001$, $df = 1$, $p > 0.999$, $\Phi_{\text{Cramer}} < 0.001$. In het databestand geeft de variabele **SEKSE** weer of de leerling een jongen of een meisje was.

Tabel 2.6*Aantal en percentage leerlingen in de populatie en de onderzoeksgroep naar sekse*

Onderdeel	Sekse		Totaal (onbekend)
	Jongen	Meisje	
Onderzoeksgroep	8222 (50.5%)	8027 (49.3%)	16297 (48)
Intelligentie	3595 (50.5%)	3500 (49.2%)	7116 (21)
Begrijpend lezen	7154 (50.2%)	7070 (49.6%)	14251 (27)
Taalverzorging	4491 (50.9%)	4317 (48.9%)	8826 (18)
Wiskunde	7098 (50.4%)	6958 (49.4%)	14085 (29)
Engels	4449 (49.9%)	4444 (49.8%)	8918 (25)
Burgerschap	2178 (49.3%)	2234 (50.6%)	4418 (6)
Leerlingvragenlijst	7250 (50.4%)	7127 (49.5%)	14394 (17)
Oudervragenlijst	3930 (49.2%)	4050 (50.7%)	7989 (9)
Populatie	102388 (50.6%)	99929 (49.4%)	202317

Representativiteit naar leeftijd

Ten slotte is de representativiteit van de onderzoeksgroep geanalyseerd vanuit de achtergrondvariabele leeftijd. In het databestand bevat de variabele **LEEFTIJD** de leeftijd van de leerlingen op 1 maart 2014. Er is gekozen om bij de constructie van de variabele *leeftijd* uit te gaan van een vaste referentiedatum om zo het aantal variabelen in het databestand enigszins beperkt te houden. Bovendien lijkt het er voor de analyses niet veel toe te doen hoe oud de leerling exact was op het tijdstip van afname. Figuur 2.1 laat voor verschillende leeftijdsgroepen het aantal leerlingen in de onderzoeksgroep zien. We zien dat de verdeling van de leeftijden enigszins rechtsscheef verdeeld is. Zoals verwacht mag worden heeft de meerderheid van de leerlingen een leeftijd die ‘past’ bij een niet vertraagde schoolcarrière. De overige leerlingen lijken ergens gedurende de schoolloopbaan een te hebben klas overgeslagen (3.3%) of vertraging te hebben opgelopen (19.2%). Onderzoek heeft laten zien dat de verhouding tussen de reguliere en vertraagde leerlingen in jaargroep 8 van het basisonderwijs rond de 80/20 ligt (zie bijvoorbeeld Heesters, van Berkel, van der Schoot, & Hemker, 2007). De COOL⁵⁻¹⁸-onderzoeksgroep lijkt qua leeftijdsopbouw dus behoorlijk representatief te zijn voor de populatie. Als we kijken naar de leeftijdsopbouw bij de afzonderlijke onderdelen zien we een vergelijkbaar beeld.



Figuur 2.1. *Leeftijdsverdeling onderzoeksgroep VO-3*

Samenvattend laten de representativiteitsanalyses zien dat de totale COOL⁵⁻¹⁸-onderzoeksgroep ($N = 16297$) voldoende representatief is voor de populatie als het gaat om de variabele *leeftijd* en *seks*. Dit betekent dat het niet nodig is om in analyses te controleren of terug te wegen voor deze variabele. Dat geldt niet voor de variabelen *onderwijstype*, *regio*, en *mate van verstedelijking*. Bij deze variabelen zien we namelijk wel dat de aantallen in de onderzoeksgroep significant afwijken van de aantallen die we mogen verwachten op grond van gegevens van DUO. Afhankelijk van de onderzoeksvraag en de analyse kan het dus wenselijk zijn om deze variabelen expliciet mee te nemen. De resultaten gelden echter alleen voor de totale onderzoeksgroep. De representativiteit van de onderzoeksgroep **binnen** een onderwijstype is nog niet in kaart gebracht. Die analyse kan eenvoudig plaatsvinden op basis van de achtergrondvariabelen die in het databestand zijn opgenomen.

3 INTELLIGENTIE

3.1 Inleiding

In het onderzoeksvoorstel van COOL⁵⁻¹⁸ wordt intelligentie opgevat als een achtergrondkenmerk, dat op meerdere momenten moet worden gemeten, aangezien "... begaafdheid geen statisch gegeven is, maar zich in complexe interactie met de omgeving ontwikkelt." (p. 6). De twee aangegeven meetmomenten zijn groep vijf van het primair onderwijs en leerjaar drie van het voortgezet onderwijs. Daarom is voor de eerste meting van COOL⁵⁻¹⁸ een extra versie van de NSCCT ontwikkeld. Een eenvoudigere versie van de NSCCT wordt namelijk binnen COOL⁵⁻¹⁸ ook in groep 5 van het primair onderwijs wordt gebruikt.

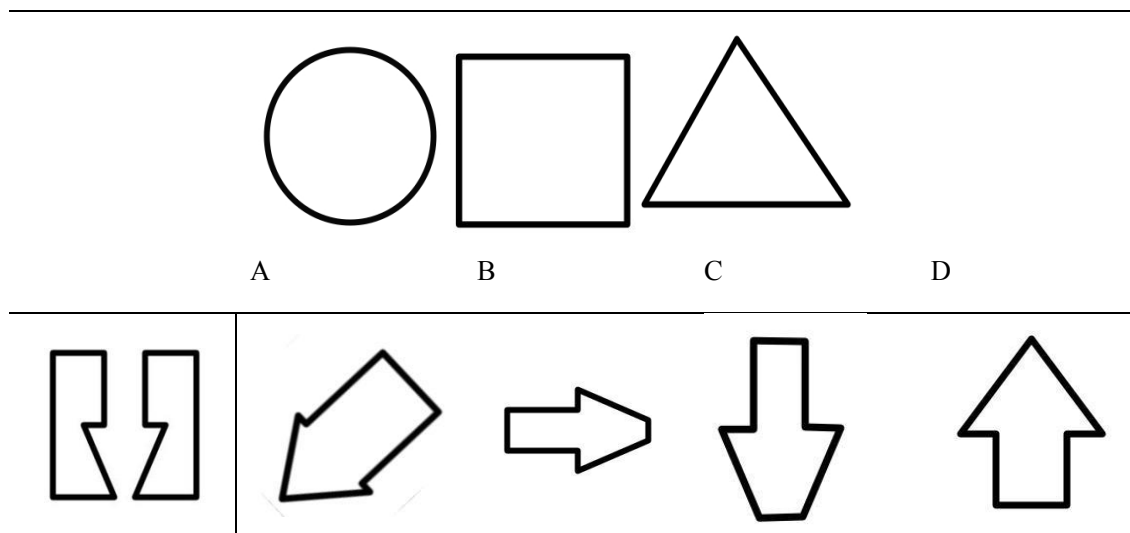
De nieuw ontwikkelde versie van de NSCCT is voorafgaand aan de meting van COOL⁵⁻¹⁸ in 2007-2008 in een proefonderzoek in de derde klas van het voortgezet onderwijs uitgetest. Op grond van de analyseresultaten is een beperkt aantal items weggelaten, waarna de nieuw ontwikkelde NSCCT-versie aan alle voorwaarden voor gebruik in het voortgezet onderwijs beantwoordde. Omdat de versie voor het voortgezet onderwijs ten dele bestaat uit de moeilijkste items van de versie voor groep acht van het primair onderwijs, is het tevens makkelijker om equivaleringsonderzoek te verrichten met de intelligentiemeting in groep vijf van het primair onderwijs. De diverse versies van de NSCCT voor het primair onderwijs zijn onderling namelijk ook 'gelinked'.

De meest geschikte verwijzing naar de NSCCT is Van Batenburg en Van der Werf (2004) – zij het dat die publicatie de versies ten behoeve van het primair onderwijs betreft.

3.2 De NSCCT-versie voor VO3

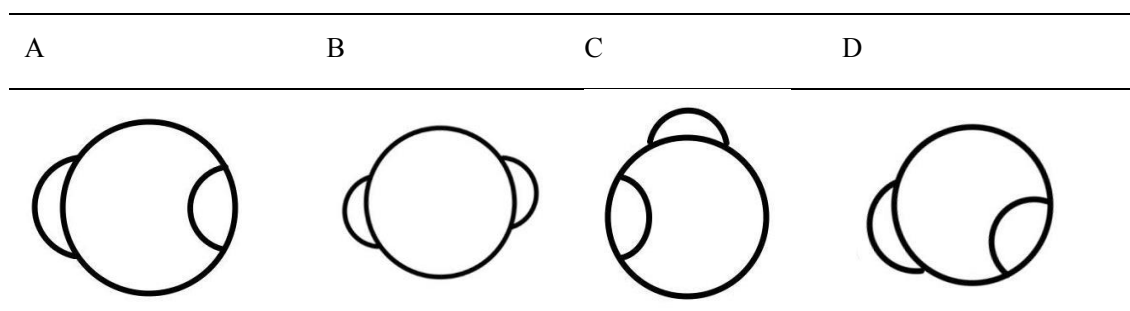
De in COOL⁵⁻¹⁸ afgenomen versie van de NSCCT bestaat uit de volgende vijf onderdelen (subtests): figuur samenstellen (17 items), exclusie (14 items), getallenreeksen (20 items), categorieën (15 items), en analogieën (10 items). Het totale aantal items is 76. Dit is exact dezelfde versie van de NSCCT die ook is afgenomen in de derde klas van het voortgezet onderwijs in 2007-2008 en een iets andere versie dan die in 2010-2011 is afgenomen. Hieronder worden voorbeeld items weergegevens van de verschillende NSCCT onderdelen.

Het onderdeel 'figuur samenstellen' refereert aan het ruimtelijk inzicht van de leerlingen. Er moet een figuur (een cirkel, vierkant of driehoek) worden samengesteld uit een gegeven deel en een complementair deel. De leerling kan kiezen uit vier antwoordalternatieven en de leerling moet zelf bepalen of de figuur een cirkel, vierkant of driehoek kan worden. Zie het voorbeeld item in Figuur 3.1.



Figuur 3.1.*Voorbeeld item van subtest figuur samenstellen*

Het onderdeel ‘exclusie’ refereert eveneens aan het ruimtelijk inzicht van leerlingen. Uit vier figuren moet er één worden gekozen die in een bepaald opzicht afwijkt van de andere drie figuren. Voor een voorbeeld van een item uit deze subtest zie Figuur 3.2.



Figuur 3.2.*Voorbeeld item van subtest exclusie*

Het onderdeel ‘getallenreeks’ refereert aan het numerieke inzicht van de leerling. Bij dit onderdeel moet op een openstaande plaats van een reeks getallen het getal worden “geplaatst” dat de reeks logisch maakt. Het gaat er dus om dat de leerling de regel ontdekt die aan de reeks ten grondslag ligt. De reeks bestaat uit minstens vijf en maximaal negen getallen, inclusief de openstaande plaats. De openstaande plaats kan zowel aan het begin, aan het eind, als ergens in het midden van de reeks zijn gepositioneerd. Er zijn vier getallen gegeven, waarvan de leerling er één moet uitkiezen (Figuur 3.3).

										A	B	C	D
0	1	2	0	2	3	-	3	4		4	2	1	0

Figuur 3.3.*Voorbeeld item van subtest getallenreeks*

Het onderdeel ‘categorieën’ refereert aan het verbale inzicht van de leerlingen. Er worden drie woorden gegeven, die op een bepaalde manier tot één categorie behoren. Uit vier andere woorden moet er één worden gekozen, dat eveneens tot die categorie behoort. Voor een voorbeeld van een item van de subtest categorieën zie Figuur 3.4.

			A	B	C	D
klei	zand	modder	water	aarde	berg	akker

Figuur 3.4.*Voorbeeld item van subtest categorieën*

Het onderdeel ‘analogieën’ refereert eveneens aan het verbale inzicht van leerlingen. Bij dit onderdeel worden twee woorden gegeven die in een bepaalde relatie tot elkaar staan. Daarachter staat een derde woord. Uit vier daarna volgende woorden moet het woord worden gekozen dat in dezelfde relatie staat tot het derde woord, als de relatie tussen de eerste twee woorden (Figuur 3.5).

				A	B	C	D
honger	- eten	dorst	-	water	drinken	kraan	woestijn

Figuur 3.4.*Voorbeeld item van subtest analogieën*

De NSCCT-opgaven stonden in een toetsboekje. Elk onderdeel begon met een specifieke instructie en een voorbeelditem. De antwoorden moesten worden aangestreept op een antwoordblad. Per subtest was een bepaalde maximale tijdsduur vastgesteld. De bij de afname aanwezige docent werd geacht de tijdsduur te bewaken. Leerlingen die binnen de vastgestelde tijdsduur al klaar waren met een onderdeel, mochten nog niet met het volgende onderdeel beginnen. De totale afnametijd bedroeg één lesuur.

Bij een dergelijke klassikale afname kan echter niet worden uitgesloten dat er in bepaalde klassen, of door bepaalde leerlingen is afgeweken van de beoogde tijdsverdeling van onderdelen. Het is mogelijk dat in bepaalde klassen de tijdsduur per onderdeel minder stringent is bewaakt dan de bedoeling was, zodat er voor bepaalde onderdelen meer tijd was dan bedoeld en voor andere onderdelen minder tijd. Tevens is het mogelijk dat bepaalde leerlingen die binnen de tijdsduur klaar waren met een onderdeel niet hebben gewacht tot ze weer verder mochten gaan, maar vooruit zijn gaan werken.

3.3 Constructie intelligentiescores

Het bestand met de ingelezen intelligentie items bevat 7116 leerlingen, die minstens één item hebben beantwoord en derhalve aan de afname van de intelligentietest hebben deelgenomen. De in deze paragraaf beschreven analyses hebben betrekking op deze 7116 leerlingen. Aan het bestand is een variabele **DUM_INT** toegevoegd die aangeeft of de intelligentietoets is gemaakt.

Het werd echter niet zinvol geacht om aan leerlingen die slechts enkele items hebben beantwoord een intelligentiescore toe te kennen (die dan uiteraard heel laag is). Bij de hierna beschreven constructie van de intelligentiescore zijn twee verschillende benaderingen gehanteerd om met veel niet beantwoorde items om te gaan. Het resultaat is dat er twee verschillende intelligentievariabelen zijn aangemaakt.

In de eerste benadering was het uitgangspunt dat minstens de helft plus één item, dat wil zeggen minstens 39 items, beantwoord moesten zijn. In de tweede benadering was het uitgangspunt dat per onderdeel minstens de helft van de items moest zijn beantwoord, dus respectievelijk 9, 7, 10, 8 en 5 items. In totaal zijn dit ook 39 items, maar toch is dit een meer stringente voorwaarde. Dat blijkt ook uit de aantallen

leerlingen die aan de twee gestelde criteria voldoen. In het eerste geval zijn dat 7089 leerlingen (99,6%); in het tweede geval 6897 leerlingen (96,9%). Per onderdeel is het aantal leerlingen dat aan het tweede criterium voldoet respectievelijk 7114, 7107, 7058, 6994, 6932. Dit afnemende aantal is een indicatie dat de tijdsduur niet in alle klassen optimaal is bewaakt. Voor alle duidelijkheid merken we op dat alle leerlingen die aan het tweede criterium voldoen, ook aan het eerste criterium voldoen. Aan het bestand zijn vijf ‘telvariabelen’ toegevoegd, die per onderdeel het aantal beantwoorde items aangeven (**INT_VAL1** tot en met **INT_VAL5**).

In beide selecties van leerlingen zijn dezelfde analyses verricht. In de eerste plaats is er per onderdeel, en overall, een betrouwbaarheidsanalyse verricht. In de tweede plaats zijn er per onderdeel somscores van het aantal goed beantwoorde items gevormd, en is op deze somscores een factoranalyse verricht. In de derde plaats is een somscore van het totaal aantal goed beantwoorde items gevormd, en is deze score in twee stappen getransformeerd naar een intelligentiescore. Ten slotte zijn er beschrijvende analyses op itemniveau, op het niveau van de vijf onderdelen en op het niveau van de totale score verricht. In paragraaf 3.3.1 en 3.3.2 vatten we de voornaamste resultaten van deze analyses samen.

3.3.1 Resultaten analyses per item en per onderdeel

In de eerste plaats zijn de *p*-waarden (moeilijkheidsgraad) van de items van belang. Tabel 3.1 toont per onderdeel de gemiddelde *p*-waarde en het minimum en maximum van de items in beide selecties. Het algemene beeld is dat er nauwelijks verschil is tussen beide selecties. Vanwege het grote aantal leerlingen en het relatief kleine verschil tussen het aantal leerlingen in de twee selecties, ligt dat ook voor de hand. Toch is er wel enig verschil, de gemiddelde *p*-waarde voor de laatste twee onderdelen is in de tweede selectie net iets hoger dan in de eerste selectie. Verder wordt duidelijk dat de gemiddelde moeilijkheid van elk onderdeel ongeveer hetzelfde is. Alleen het laatste onderdeel is iets moeilijker dan de voorgaande vier onderdelen.

Naast de gemiddelde *p*-waarde is ook de spreiding van de *p*-waarden van belang. Voor de eerste selectie was de spreiding van *p*-waarden over de test als geheel als volgt: [0 - 0.39] 5 items, [.40 - .49] 9 items, [.50 - .59] 13 items, [.60 - .69] 14 items, [.70 - .79] 13 items, [.80 - .89] 18 items, [.90 - 1] 4 items. We kunnen concluderen dat er een goede spreiding van de *p*-waarden is. De spreiding van *p*-waarden voor de tweede selectie was vergelijkbaar met de spreiding van *p*-waarden voor de eerste selectie.

Tabel 3.1

Samenvatting p-waarden per onderdeel in beide selecties

Deel	Selectie criterium 1 (N = 7089)			Selectie criterium 2 (N = 6897)		
	<i>M</i>	Min.	Max.	<i>M</i>	Min.	Max.
1	.66	.29	.95	.66	.29	.95
2	.64	.42	.95	.64	.42	.95
3	.68	.42	.88	.68	.42	.88
4	.66	.21	.86	.67	.21	.87
5	.61	.23	.89	.63	.23	.91

Vervolgens is er in beide selecties per onderdeel een betrouwbaarheidsanalyse verricht. De resultaten daarvan zijn samengevat in Tabel 3.2. Per selectie laat de eerste kolom de gevonden betrouwbaarheid

(Cronbach's α , of ook wel λ_3) zien. Omdat de betrouwbaarheid ten dele afhangt van het aantal items, dat tussen de onderdelen verschilt, laten we in de tweede kolom van beide selecties de gemiddelde inter-item correlatie zien (de tweede 'component' die bepalend is voor de betrouwbaarheid). Dit gemiddelde geeft zoiets als de 'homogeniteit' van de itemset aan. De derde kolom laat in beide gevallen de betrouwbaarheid zien, die verkregen kan worden door één relatief laag (of eventueel zelfs negatief) correlerend item weg te laten.

Voor items binnen het cognitieve domein zijn de gemiddelde inter-item correlaties normaal te achten, en de betrouwbaarheidscoëfficiënten per onderdeel – het aantal items in aanmerking genomen – acceptabel. Wat vooral opvalt is dat de betrouwbaarheid in de tweede selectie iets lager is voor de twee laatste onderdelen. Hieruit moet echter niet worden geconcludeerd dat de tweede selectie "slechter" is. Met name de hogere betrouwbaarheid voor het vijfde onderdeel in de eerste selectie is ten dele toe te schrijven aan het feit dat er binnen deze selectie 146 leerlingen zijn die op geen enkel item een antwoord hebben gegeven (hetzij goed, hetzij fout). Op het vierde onderdeel zijn dat er binnen deze selectie 74, op het derde onderdeel 6, en op de eerste twee onderdelen betreft dit 0 leerlingen. Naarmate het aantal leerlingen dat geen enkel item heeft beantwoord, maar wel meetelt, en dus geacht wordt alle items 'fout' te hebben beantwoord hoger is, neemt de correlatie tussen de items, en daarmee de betrouwbaarheid toe. Verder is er op vier van de vijf onderdelen één item, zonder welk item de betrouwbaarheid van de overige items hoger zou zijn. Over het algemeen gaat het om een kleine betrouwbaarheidswinst. In de bepaling van de intelligentiescores zijn deze items gehandhaafd.

Tabel 3.2

Resultaten betrouwbaarheidsanalyse per onderdeel in beide selecties

Deel	Selectie criterium 1 ($N = 7089$)			Selectie criterium 2 ($N = 6897$)		
	α	Gem. r_{ii}	α^+	α	Gem. r_{ii}	α^+
1	.72	.14	.74	.72	.14	.74
2	.65	.12	.66	.66	.12	.66
3	.85	.22	.85	.85	.22	.85
4	.77	.19	.79	.74	.17	.76
5	.59	.13	.62	.50	.10	.55

Na de betrouwbaarheidsanalyse zijn per onderdeel de somscores bepaald, waarbij elk goed antwoord voor 1 punt telde, en elk fout of niet gegeven antwoord voor 0 punten. Deze somscores zijn aan het bestand toegevoegd, met als namen **INT_SOM1** tot en met **INT_SOM5**. Tabel 3.3 laat de correlaties tussen deze somscores zien. In het gedeelte boven de diagonaal staan de correlaties die verkregen worden binnen de eerste selectie, in het gedeelte onder de diagonaal de correlaties binnen de tweede selectie.

In de meeste gevallen is de correlatie in de tweede selectie iets hoger dan de overeenkomstige correlatie in de eerste selectie. Alleen voor de correlatie tussen het vierde en vijfde onderdeel geldt het omgekeerde. Hierbij speelt het een rol dat in de eerste selectie een aantal leerlingen van beide onderdelen geen enkel item heeft beantwoord. De correlaties tussen de vijf somscores zijn hoger dan in de NSCCT-versies voor groepen 4, 6 en 8 van het basisonderwijs (Van Batenburg & Van der Werf, 2004) en iets hoger dan in de COOL⁵⁻¹⁸ afname in 2007-2008 waarin dezelfde versie van de NSCCT is afgenomen (Zijsling et al., 2009).

Tabel 3.3*Correlaties tussen de somscores op de vijf onderdelen.*

	INT_SOM1	INT_SOM2	INT_SOM3	INT_SOM4	INT_SOM5
INT_SOM1		.57	.51	.41	.37
INT_SOM2	.57		.55	.41	.38
INT_SOM3	.52	.55		.46	.44
INT_SOM4	.43	.43	.48		.52
INT_SOM5	.41	.43	.48	.47	

Een factoranalyse op de vijf somscores laat duidelijk zien dat er slechts één onderliggende factor is. In de eerste selectie is de eerste eigenwaarde 2.86 (57.14% gemeenschappelijke variantie), en de tweede 0.77. In de tweede selectie is de eerste eigenwaarde 2.92 (58.29% gemeenschappelijke variantie), en de tweede .67. De ladingen van de vijf somscores op de eerste (enige) principale component zijn in de eerste selectie respectievelijk .76, .78, .79, .74 en .71. In de tweede selectie zijn de ladingen respectievelijk .77, .79, .80, .73 en .72. Elk onderdeel van de NSCCT behoort dus in vrijwel dezelfde mate tot de onderliggende intelligentiefactor.

3.3.2 Resultaten overall analyse

Na de analyses op het niveau van de afzonderlijke items en onderdelen, is de betrouwbaarheid van de gehele test bepaald. Hiervoor is de overall betrouwbaarheid van de gehele itemset bepaald. Aangezien de vijf onderdelen één gemeenschappelijke factor vormen, is dat een redelijke benadering. In beide selecties is de overall α gelijk aan .91. De betrouwbaarheid van de gecombineerde intelligentietest is dus alleszins acceptabel. Tabel 3.4 geeft de voornaamste statistics van de somscore, genomen over alle 76 items, binnen beide selecties.

Tabel 3.4*Karakteristieken somscore*

	Selectie criterium 1 (<i>N</i> = 7089)	Selectie criterium 2 (<i>N</i> = 6897)
Minimum	10	10
Maximum	73	73
Gemiddelde	49.84	50.20
Standaardfout	0.14	0.14
Standaarddeviatie	11.94	11.77
Skewness	-0.52	-0.54
Kurtosis	-0.25	-0.18

De negatieve waarde van de ‘skewness’ wijst erop dat de verdeling enigszins scheef naar rechts is, en de negatieve waarde van de ‘kurtosis’ wijst op een enigszins platte verdeling, ten opzichte van een standaard normale verdeling. Aangezien de theoretische verdeling van intelligentie wel ‘standaard normaal’ is, is een transformatie toegepast volgens de procedure van Blom (1958). De met deze transformatie verkregen scores zijn vervolgens via een lineaire transformatie omgezet in intelligentiescores met een gemiddelde van 100 en een standaard deviatie van 15. Dit is in beide selecties gedaan, zodat twee enigszins verschillende

intelligentiescores zijn verkregen. Deze zijn aan het bestand toegevoegd met de namen **INT_SCO1** en **INT_SCO2**. Ter controle toont Tabel 3.5 de karakteristieken van **INT_SCO1** en **INT_SCO2**.

In beide selecties is de verdeling vrijwel standaard normaal geworden. Voor een nader begrip van beide variabelen is het nuttig te vermelden dat de correlatie tussen beide variabelen, binnen de groep waarin beide variabelen gedefinieerd zijn (dat is de tweede selectie), precies 1.00 is. Het is echter niet zo dat in deze groep alle scores hetzelfde zijn. Het verschil **INT_SCO1** minus **INT_SCO2** is voor de meeste leerlingen positief, met een range van -.12 tot .77.

Verder dienen we op te merken dat het omzetten naar scores met een gemiddelde van 100 en een standaard deviatie van 15 plaats heeft gevonden binnen de verkregen steekproef, die geen representatieve afspiegeling van de totale populatie is. Immers, de leerlingen met de in de totale populatie laagste intelligentiescores kunnen niet in de derde klas van het regulier voortgezet onderwijs zijn terecht gekomen. Andersom bekeken kan het uitgesloten worden geacht dat leerlingen in de derde klas van het regulier voortgezet onderwijs een IQ onder de 50 hebben. Dat deze scores toch voorkomen is louter een gevolg van de toegepaste procedure.

Het omzetten naar scores met een gemiddelde van 100 en een standaard deviatie van 15 betekent ook dat de scores die in dit rapport beschreven worden alleen gebruikt kunnen worden om analyses te verrichten *binnen* de derde meting van COOL⁵⁻¹⁸ in VO-3. Vergelijkingen met de eerste twee metingen van COOL⁵⁻¹⁸ kunnen *niet* gemaakt worden op basis van dit rapport en het bijbehorende databestand.

Tabel 3.5

Karakteristieken intelligentievariabelen

	INT_SCO1	INT_SCO2
Minimum	43.74	43.84
Maximum	149.89	149.77
Gemiddelde	99.99	99.99
Standaardfout	0.18	0.18
Standaarddeviatie	14.97	14.97
Skewness	-0.01	-0.01
Kurtosis	-0.03	-0.03

4 BEGRIJPEND LEZEN

4.1 Materiaal

Bij de meting in 2014 zijn dezelfde toetsen voor begrijpend lezen gebruikt als in 2011. In Zijsling et al. (2013) is een uitgebreide beschrijving van de inhoud van de toetsen te vinden. Tijdens de toetsafname werden vragen gesteld over leesteksten. Bij het beantwoorden van de vragen konden leerlingen steeds kiezen uit drie of vier antwoordalternatieven. Hieronder staan enkele voorbeeldvragen.

Samenvatten: **De slimme tabaksfabrikant**

Welk kopje geeft het beste de inhoud weer van deel 3 (alineas 5 en 6)

- A Mannen, stoere rokers
- B Reclame, niet gek genoeg
- C Roken, goed voor de rust
- D Tabaksreclame lang ongehinderd

Interpreteren: **De slimme tabaksfabrikant**

Wat is het verband tussen alinea 3 en 4?

- A Alinea 3 en 4 vormen een tegenstelling.
- B Alinea 4 snijdt een nieuw onderwerp aan.
- C Alinea 4 trekt een conclusie uit alinea 3.
- D Alinea 4 werkt de informatie uit alinea 3 verder uit.

Begrijpen: **De slimme tabaksfabrikant**

Tabaksfabrikanten probeerden met reclame mensen te overtuigen dat roken geen probleem was. Hoeveel positieve eigenschappen van roken worden er in alinea's 5 en 6 genoemd?

- A 2 eigenschappen
- B 3 eigenschappen
- C 4 eigenschappen
- D 5 eigenschappen

De bijbehorende tekst is terug te vinden in Bijlage 1.

4.2 Onderzoeksdesign

Tijdens de meting in 2014 zijn in totaal 78 items afgenomen. De items zijn afgenomen volgens een structureel onvolledig design. In geval niet alle items bij alle leerlingen in de onderzoeksgroep worden afgenomen, kunnen de items alleen op dezelfde meetschaal gebracht worden indien de deeldesigns iets gemeenschappelijk hebben. De gemeenschappelijkheid kan liggen in de leerlingen die verschillende items maken, dan wel in de items die door verschillende leerlingen worden gemaakt (Eggen, 1993). Figuur 4.1 geeft het design zoals dat gebruikt is tijdens de afname.

Eerst zijn de items op basis van een inschatting van hun moeilijkheidsgraad verdeeld in vier modules (A tot en met D) met een wisselend aantal items: module A bevatte 18 items, module B 21 items, module C 18 items, en module D 21 items. Vervolgens zijn per onderwijstype twee modules geselecteerd die qua inhoud en moeilijkheidsgraad geschikt leken voor afname. VMBO leerlingen kregen afhankelijk van de leerweg modules A, B, of C voorgelegd en HAVO en VWO leerlingen kregen modules C en D voorgelegd. In het design is er voor gezorgd dat er zogenaamde ankermodules waren. Dat betekent dat een bepaalde groep leerlingen steeds een aantal items gemeenschappelijk had met een andere groep leerlingen. Vanwege de ankermodules is het mogelijk om de leesitems in de analyse op één onderliggende meetschaal te plaatsen (zie bijvoorbeeld Kolen & Brennan, 1995; Eggen, 2004). In het databestand geeft de variabele **BGL_VERS** aan welke combinatie van modules aan de leerling is voorgelegd.

Versie	Onderwijstype	Module			
		A	B	C	D
8	BBL				
9	KBL/GL/TL				
10	HAVO/VWO				

Figuur 4.1. *Afnamedesign begrijpend lezen meting 2*

Een deel van de items die bij de metingen in 2011 en 2014 is gebruikt, is ook al bij eerdere metingen gebruikt, namelijk bij de eerste meting van COOL⁵⁻¹⁸ in 2008 (zie Zijssling et al., 2009) en bij de laatste meting van VOCL'99 in 2002 (zie Kuyper & Van der Werf, 2005; Zijssling, Kuyper, Lubbers & Van der Werf, 2005). Door die overlap ontstaan er mogelijkheden om de verschillende metingen in de tijd aan elkaar te relateren. Deze koppeling wordt in een aanvullend rapport beschreven (Keuning, Keizer-Mittelhaeuser, & Timmermans, in voorbereiding). Dit betekent dat de scores die in dit rapport beschreven worden alleen gebruikt kunnen worden om analyses te verrichten *binnen* de derde meting van COOL⁵⁻¹⁸ in VO-3. Vergelijkingen met de eerste twee metingen van COOL⁵⁻¹⁸ of de laatste meting van VOCL'99 kunnen *niet* gemaakt worden op basis van dit rapport en het bijbehorende databestand.

4.3 Meetmodel

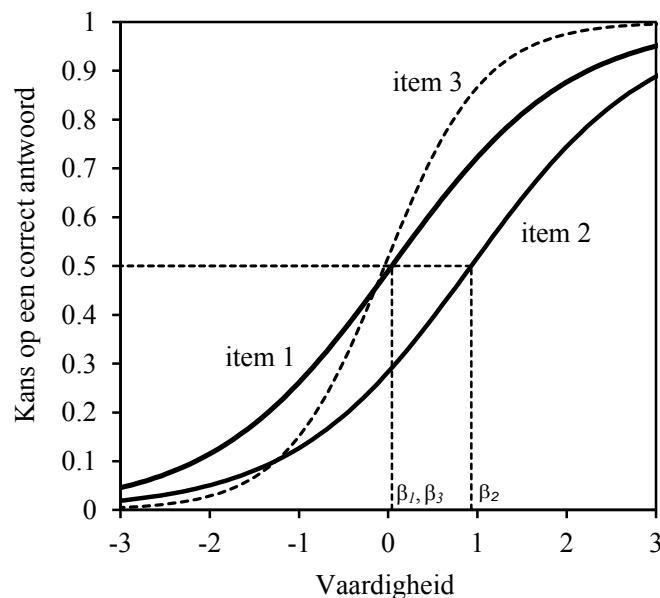
Er is bij de analyse gebruikgemaakt van een meetmodel uit de item respons theorie. Item respons modellen verschillen in een aantal opzichten nogal sterk van de modellen uit de klassieke testtheorie (Verhelst, 1993). Binnen de item respons theorie staat de latente vaardigheid centraal in plaats van de ware score op een specifieke toets. Er wordt verondersteld dat de vaardigheid θ van een persoon niet begrensd is en dat de antwoorden op bepaalde items informatie geven over de vaardigheid van de persoon. De relatie tussen de latente vaardigheid en de itemantwoorden is gedefinieerd in een zogenaamde item karakteristieke functie. De item karakteristieke functie drukt de kans op het correct beantwoorden van item j uit als functie van θ . De kenmerken en het verloop van de item karakteristieke functie zijn gespecificeerd in een aantal randvoorwaarden: (a) de kans op een correct antwoord moet tussen 0 en 1 liggen, (b) de functie moet continu zijn, en (c) de functie moet strikt stijgend zijn binnen θ . Hoewel een groot aantal wiskundige functies uitgesloten wordt door deze drie vereisten, blijven er nog veel functies over die aan de gestelde eisen voldoen. Door één specifieke functie te kiezen, kan de theorie ingeperkt worden tot één speciaal 'geval', een zogenaamd item respons model.

De antwoorden van de leerlingen op de items voor begrijpend lezen zijn geanalyseerd met het One-Parameter Logistic Model (OPLM) van Verhelst en Glas (1995). De item respons functie voor het OPLM wordt gegeven door de volgende vergelijking:

$$P(X_j = 1 | \theta) = P_j(\theta) = \frac{\exp[a_j(\theta - \beta_j)]}{1 + \exp[a_j(\theta - \beta_j)]}, \quad \text{voor } j = 1, \dots, k$$

waarin $P_j(\theta)$ de kans is dat een random gekozen persoon met vaardigheid θ item j correct beantwoordt, a_j en β_j getallen zijn die item j karakteriseren, k het aantal items in de toets is, en waar 'exp' een constante is met waarde 2.718. Er kan nagegaan worden dat functie $P_j(\theta)$ altijd tussen 0 en 1 ligt en dat $P_j(\theta)$ gelijk is aan 0.50 als de vaardigheid gelijk is aan β_j . De waarde voor β_j kan daarom ook geïnterpreteerd worden als de hoeveelheid vaardigheid die vereist is om een kans van 50 procent te hebben op het correct maken van een item.

Figuur 4.2 geeft drie verschillende item respons curven voor het OPLM. Zoals we kunnen zien, is voor item 2 meer vaardigheid nodig om die kans van 50 procent te bereiken dan voor item 1 en 3. In feite reflecteert parameter β_j dus de moeilijkheid voor item j of, meer formeel, de positie van de item karakteristieke curve in relatie tot de vaardigheidsschaal. Om deze reden wordt β_j in de literatuur vaak aangeduid als de moeilijkheids- of locatieparameter van item j . Figuur 4.2 laat verder zien de item respons curven voor item 1 en item 2 qua vorm hetzelfde zijn en dat de vorm van item 3 afwijkt. De afwijkende vorm van de curve voor item 3 wordt veroorzaakt door een andere waarde voor a_j . Er kan afgeleid worden dat item 3 in een bepaald vaardigheidsgebied beter discrimineert tussen personen, omdat de kans op het correct maken van het item snel toeneemt naarmate de vaardigheid hoger is. Bij item 1 en 2 neemt de kans op het correct maken van het item minder snel toe als de vaardigheid stijgt. Item 1 en 2 discrimineren in dat vaardigheidsgebied dus minder goed tussen personen.



Figuur 4.2. Drie itemresponscurven voor het OPLM

Kenmerkend voor het OPLM is dat a_j niet uit de data geschat wordt, maar a priori als constante wordt ingevoerd. Door deze keuze kan de (gewogen) ruwe score direct uit de data berekend worden en behoort toepassing van de *conditional maximum likelihood* methode voor het schatten van de itemparameters tot de mogelijkheden. Een belangrijk voordeel van *conditional maximum likelihood* is dat er bij het schatten van de itemparameters geen veronderstelling hoeft te worden gedaan over de verdeling van de vaardigheid in de populatie. Bovendien doet het er niet toe hoe de steekproef getrokken is of hoe de modules in een incompleet design zijn samengesteld. Zelfs als een leerling meerdere keren meedoet aan het onderzoek of bepaalde modules systematisch bij bepaalde groepen leerlingen worden afgenomen, kunnen de itemparameters adequaat geschat worden (Eggen, 2004).

De schatting van de itemparameters vindt plaats met behulp van het computerprogramma OPLM (Verhelst, Glas, & Verstralen, 1995). Dit programma voert tevens een aantal statistische toetsen uit op grond waarvan bepaald kan worden of het model een adequate beschrijving geeft van de data. Belangrijk zijn de zogenaamde itemgeoriënteerde *S*-toets en de overall *R**I**c*-toets. De *S*-toets is asymptotisch χ^2 verdeeld en is gebaseerd op de verschillen tussen de geobserveerde en verwachte proporties antwoorden in homogene scoregroepen. Een uniforme verdeling van *p*-waarden voor de *S*-toetsen in het interval [0,1] pleit voor passing van het model (zie Verhelst, Glas, & Verstralen, 1995). De *R**I**c*-toets heeft dezelfde onderliggende rationale als de *S*-toets en wordt over het algemeen acceptabel bevonden indien zijn waarde niet groter is dan anderhalf tot hooguit twee keer het aantal vrijheidsgraden.

4.4 Kalibratie

Er zijn tijdens de meting in 2014 drie verschillende versies voor begrijpend lezen gebruikt (zie Figuur 4.1). Doordat niet alle leerlingen in de onderzoeksgroep dezelfde toets gemaakt hebben, zijn de toetsresultaten van verschillende leerlingen niet altijd met elkaar te vergelijken. Om dit probleem te ondervangen zijn de toetsen gekalibreerd met behulp van het OPLM. Als het OPLM geldt voor een verzameling items kan de vaardigheid van een leerling geschat worden met elke willekeurige deelverzameling items. De (gewogen) scores van de leerlingen kunnen dus vertaald worden naar zogenaamde vaardigheidsscores die over versies van toetsen vergelijkbaar zijn.

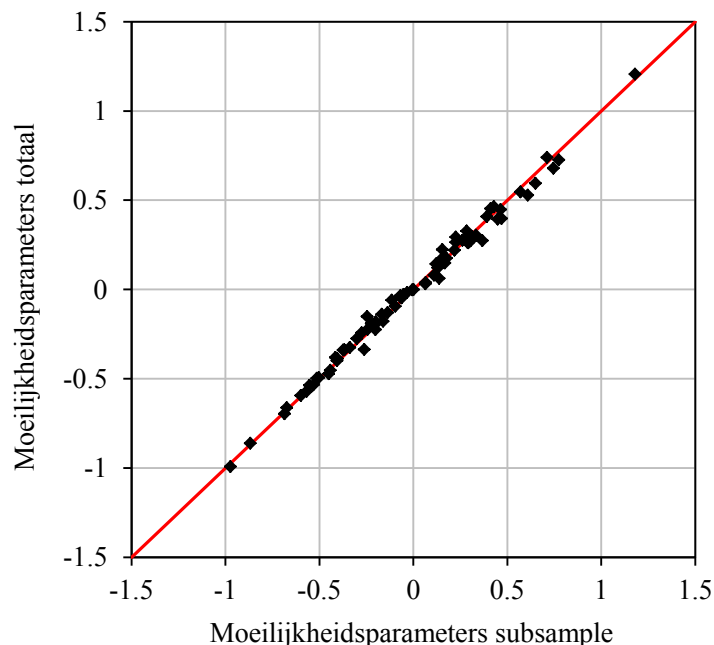
Een kalibratie met het OPLM omvat een aantal stappen. Na de specificatie van de items en het design worden de itemgewichten, a_j , ingesteld in het computerprogramma OPLM. Vervolgens worden de moeilijkheidsparameters, β_j , geschat met behulp van de *conditional maximum likelihood* methode. Aansluitend wordt voor elk item de passing beoordeeld op basis van de beschikbare statistische toetsen. Tevens vindt er een grafische inspectie van de item karakteristieke curven plaats met het programma OPDRAW. Ten slotte vindt er een globale modelcontrole plaats middels de *R**I**c*-toets en de verdeling van de overschrijdingskansen van de *S*-toetsen. De verschillende stappen worden in de regel een aantal malen doorlopen totdat het resultaat bevredigend is. Afhankelijk van de uitkomsten kunnen items worden verwijderd.

Als gevolg van het afnamedesign (zie Figuur 4.1) varieerde het aantal waarnemingen per item sterk. De items in module A zijn bijvoorbeeld door 537 leerlingen gemaakt, terwijl de items in module C door 10283 leerlingen zijn gemaakt. In principe doet het er voor het schatten van het OPLM niet toe hoeveel leerlingen een item hebben gemaakt zolang er maar voldaan wordt aan een zeker minimum. Voor een adequate schatting van het OPLM is het een voorwaarde dat er ten minste 300 tot 400 responsen beschikbaar zijn per item (Keuning, 2004; Evers, Lucassen, Meijer & Sijsma, 2010). Aan deze eis wordt ruimschoots voldaan. Door het ongelijke aantal waarnemingen per item is het echter problematisch om de passing van het model

te beoordelen. Statistische toetsen zijn immers in hoge mate gevoelig voor het aantal waarnemingen. Bij een groot aantal waarnemingen is een modelafwijking al snel significant, terwijl bij een klein aantal waarnemingen haast geen enkele modelafwijking significant zal zijn. Om dergelijke problemen bij het beoordelen van de passing van het model te vermijden, is de analyse in twee stappen uitgevoerd. In de eerste stap is het OPLM geschat op basis van alle beschikbare waarnemingen. In een vervolgstap is het OPLM geschat op basis van een random *sample* uit de totale dataset waarin het aantal waarnemingen per item varieerde van minimaal 400 tot maximaal 800. Bij de beoordeling van de passing van het OPLM is in eerste instantie uitgegaan van de resultaten zoals die verkregen werden op basis van de *subsample*. Daarna is de samenhang tussen de schattingen op basis van de totale dataset en de schattingen op basis van de *subsample* in kaart gebracht. Als het model past zou de relatie tussen beide sets met itemparameterschattingen lineair moeten zijn (cf. Lord & Novick, 1968; Hambleton, Swaminathan, & Rogers, 1991). De uiteindelijke kalibratie is gebaseerd op de dataset met alle waarnemingen.

De kalibratie op basis van de *subsample* liet zien dat de prestaties van de leerlingen op de leesitems adequaat beschreven kunnen worden door het OPLM. Ten eerste bleek de verdeling van p -waarden (overschrijdingskansen) voor de S -toetsen redelijk uniform verdeeld te zijn. Ten tweede was de verhouding tussen de R^2 -bijdrage en het aantal vrijheidsgraden acceptabel, $R^2 = 593$, $df = 362$, $p < .01$. De kalibratie op basis van de *subsample*, ten slotte, vertoonde grote gelijkenis met de kalibratie op basis van de dataset met alle waarnemingen. Figuur 4.3 laat duidelijk zien dat de relatie tussen de twee sets met moeilijkheidsparameters lineair is.

Bij drie leesitems bleken de prestaties van leerlingen niet goed beschreven te kunnen worden door het OPLM (nummer 34 en 31 in versie 9; nummer 19, 13 en 38 in versie 10). Deze items zijn daarom uit de meetschaal gelaten.



Figuur 4.3. Relatie tussen moeilijkheidsparameters in twee verschillende kalibraties

4.5 Meetnauwkeurigheid

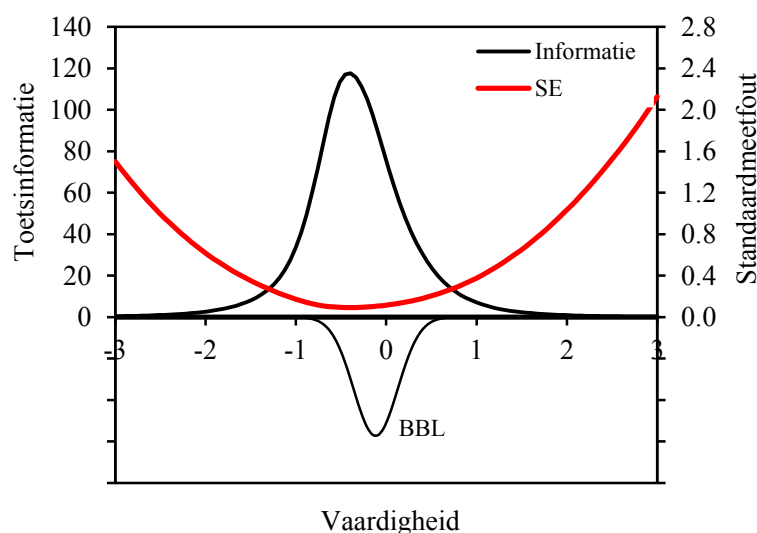
Binnen de item respons theorie wordt gebruikgemaakt van het statistische begrip informatie om de meetnauwkeurigheid van een toets lokaal uit te drukken (zie bijvoorbeeld Lindgren, 1976). De iteminformatiefunctie geeft aan welke bijdrage een item levert aan de meetnauwkeurigheid voor ieder punt op de vaardigheidsschaal waar het item betrekking op heeft. De iteminformatiefunctie van een dichotoom item bereikt zijn maximum wanneer $\theta = \beta_j$ en wordt kleiner naarmate het verschil tussen θ en β_j toeneemt. Met behulp van de informatiefuncties van de afzonderlijke items in een willekeurige toets t kan een zogenaamde toetsinformatiefunctie bepaald worden. Voor de toetsinformatiefunctie $I_t(\theta)$ geldt:

$$I_t(\theta) = \sum_{j=1}^{k_t} I_j(\theta)$$

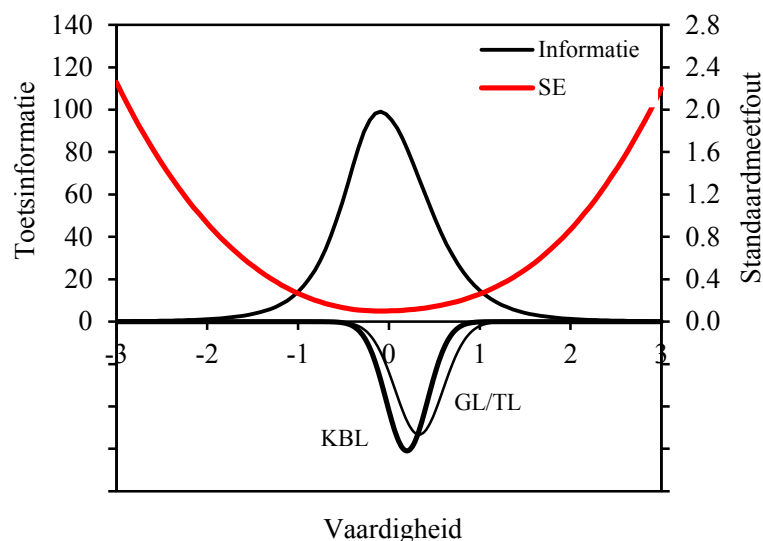
De waarde van $I_t(\theta)$ hangt in de regel samen met het aantal items in de toets, de kwaliteit van de items, en de mate van overeenstemming tussen de moeilijkheidsgraad van de items en de vaardigheid van de leerling. Uit de toetsinformatiefunctie kan de standaardmeetfout van een individuele vaardigheidsschatting afgeleid worden:

$$SE_t(\hat{\theta}) = 1/\sqrt{I_t(\theta)}$$

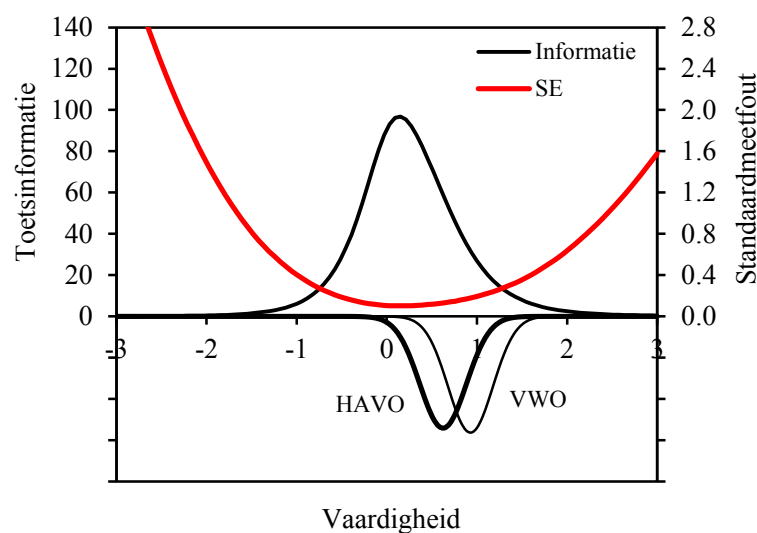
De interpretatie van deze standaardmeetfout is hetzelfde als in de klassieke testtheorie. Figuren 4.4a, 4.4b en 4.4c geven voor elke toetsversie (zie Figuur 4.1 en **BGL_VERS** in databestand) de toetsinformatiefunctie en de standaardmeetfout. Onder de x -as staan de kansdichtheidsfuncties (de ongewogen vaardigheidsverdelingen zoals geschat via *marginal maximum likelihood* met gefixeerde itemparameters) van de groepen leerlingen die de toets gemaakt hebben. De figuren laten zien dat de standaardmeetfout duidelijk het kleinst is in het vaardigheidsgebied van de doelgroep. Alleen voor VWO leerlingen, en in iets mindere mate voor HAVO-leerlingen, lijkt de toets iets te gemakkelijk te zijn geweest.



Figuur 4.4a. Toetsinformatiefunctie begrijpend leestoets versie 8



Figuur 4.4b. *Toetsinformatiefunctie begrijpend leestoets versie 9*



Figuur 4.4c. *Toetsinformatiefunctie begrijpend leestoets versie 10*

Binnen het raamwerk van de item respons theorie is het enigszins problematisch om de meetnauwkeurigheid te interpreteren, omdat er geen ondubbelzinnige criteria beschikbaar zijn om de toetsinformatiefunctie te beoordelen. De definitie uit de klassieke testtheorie van meetnauwkeurigheid is gemakkelijker te interpreteren, maar niet direct vergelijkbaar met de lokale maat voor meetnauwkeurigheid uit de item respons theorie. Wanneer we de variatie in standaardmeetfouten negeren en de variantie van de vaardigheidsverdeling vastzetten, wordt het echter mogelijk om de item respons theorie maat voor meetnauwkeurigheid te relateren aan de klassieke maat voor meetnauwkeurigheid. Verhelst, Glas, en Verstralen (1995) hebben in dit kader de *Macc* coëfficiënt ontwikkeld. De *Macc* coëfficiënt kan opgevat worden als een generalisatie van coëfficiënt alpha voor alle schattingen van θ . Omdat er binnen COOL⁵⁻¹⁸ gebruikgemaakt is van het principe niveau gestratificeerd toetsen is de *Macc* coëfficiënt eerst uitgerekend in de totale populatie voor alle items (er is net gedaan of alle items in de itembank zijn afgenomen bij alle leerlingen). Vervolgens is met behulp van de Spearman-Brown-formule voor testverlenging (of hier – verkorting) (Brown, 1910; Spearman, 1910) een schatting gemaakt van de betrouwbaarheid in geval k items zouden zijn afgenomen. De *Macc* coëfficiënt in de totale populatie voor alle $k_0 = 75$ geschaalde items bleek gelijk te zijn aan .91. De betrouwbaarheid voor $k_t = 39$ is dan gelijk aan:

$$\rho_t = \frac{\frac{k_t}{k_0} \times \rho_0}{1 + \left(\frac{k_t}{k_0} - 1\right) \times \rho_0} = \frac{\frac{39}{75} \times .91}{1 + \left(\frac{39}{75} - 1\right) \times .91} = .84$$

Voor $k_t = 37$ (versie 9) en $k_t = 36$ (versie 10) is de betrouwbaarheid afgerond gelijk aan .83. De hiervoor genoemde toetslengtes wijken af van de toetslengtes die volgen uit Figuur 4.1, omdat enkele items niet in de meetschaal bleken te passen. De betrouwbaarheidsschattingen die verkregen worden na toepassing van de Spearman-Brown-formule zijn bovendien een onderschatting van de werkelijke betrouwbaarheid. Een toetsversie bestaat immers niet uit een *random* selectie van items (de aanname waaronder de betrouwbaarheid is berekend), maar uit een selectie die *a priori* op basis van informatie over de moeilijkheidsgraad en het discriminerend vermogen van de items is gemaakt.

4.6 Toetsresultaten

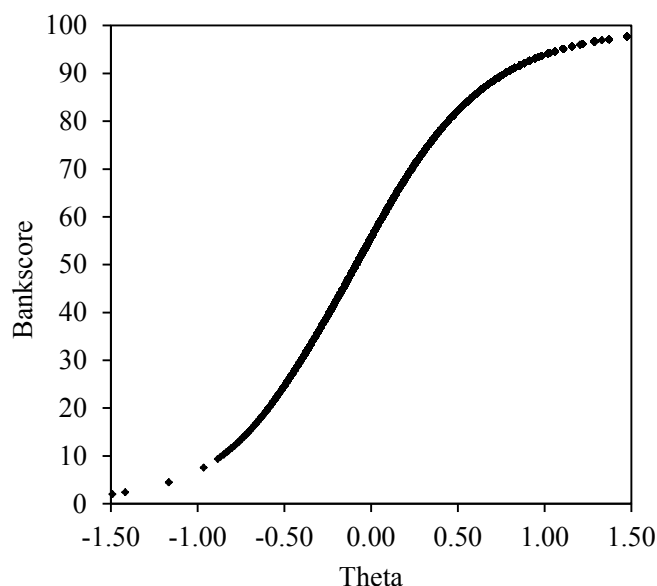
Als het OPLM geldt voor een itemverzameling kan de vaardigheid van leerlingen met elke willekeurige deelverzameling op dezelfde meetschaal geschat worden. De (gewogen) scores van de leerlingen op de drie toetsversies kunnen dus vertaald worden naar zogenaamde vaardigheidsscores die zonder problemen met elkaar vergeleken kunnen worden, ook als verschillende toetsen zijn afgenomen. Tevens is het mogelijk om op basis van de vaardigheidsscore van een leerling een zogenaamde bankscore te bepalen (zie bijvoorbeeld Hambleton, Swaminathan, & Rogers, 1991). Een belangrijk voordeel van bankscores is dat de scores geïnterpreteerd kunnen worden als een beheersingsniveau. De bankscores liggen namelijk tussen 0 en 100 procent. Bij een bankscore van 70 kun je dus zeggen dat de leerling naar verwachting 70 procent van de items in de set correct kan maken.

In het databestand staan ruwe scores, gewogen scores, vaardigheidsscores en bankscores. De betreffende variabelen heten respectievelijk **BGL_RSCO**, **BGL_WSCO**, **BGL_THET** en **BGL_BANK3**. Bij het bepalen van de ruwe en gewogen scores zijn alleen de geschaalde items meegenomen. Dit betekent dat de maximale ruwe score voor versie 8 gelijk is aan 39, voor versie 9 gelijk aan 37 en voor versie 10 gelijk aan 36. De maximale gewogen scores zijn respectievelijk gelijk aan 135, 123 en 119. De bankscores zijn niet-lineaire transformaties van de vaardigheidsscores en kunnen worden uitgerekend via de itemparameters die verkregen zijn na schatting van het OPLM:

$$\pi(\theta) = \frac{1}{\sum_{j=1}^k \alpha_j} \sum_{j=1}^k \alpha_j p_j(\theta)$$

Figuur 4.5 geeft een visuele weergave van de relatie tussen de vaardigheidsscore θ en de bankscore π . We zien dat relatie tussen de vaardigheidsscores en de bankscores in het midden nagenoeg lineair is en aan de uiteinden van de meetschaal vlakker. Dit komt doordat leerlingen met een hoge vaardigheidsscore vanaf een bepaald vaardigheidsniveau (vrijwel) alle items correct zullen maken en leerlingen met een lage vaardigheidsscore (vrijwel) alle items fout. Bij het doen van analyses op populatieniveau heeft het de voorkeur om de bankscores uit het databestand te gebruiken, omdat het gebruik van de vaardigheidsscores kan leiden tot een (sterke) overschatting van de variantie in de populatie (cf. Wu, 2005). Bij het gebruik van

bankscores speelt dit probleem niet. Voor analyses binnen een onderwijstype (of toetsversie) heeft het de voorkeur om de ruwe toetsscore te gebruiken.



Figuur 4.5. *Visuele weergave van de relatie tussen vaardigheidsscores en bankscores*

4.7 Aanvullend onderzoek in Limburg

De doelstelling van COOL⁵⁻¹⁸ is om bij een landelijk representatieve groep met leerlingen gegevens te verzamelen over burgerschap, cognitieve capaciteit, sociaal-emotionele ontwikkeling en de schoolse vaardigheden begrijpend lezen, taalverzorging, wiskunde en Engels. Daarnaast wordt via een oudervragenlijst achtergrondinformatie verzameld. In de provincie Limburg liep in 2014 een vergelijkbaar onderzoek met de naam Inventaar 3VO. Om de scholen in Limburg niet te belasten met twee in opzet vergelijkbare onderzoeken kende COOL⁵⁻¹⁸ in Limburg een andere opzet dan in de rest van Nederland. In overleg met de onderzoekers van Inventaar 3VO is namelijk een gecombineerde afname opgezet voor een achttal scholen. In deze afname werd het tekstbegrip van leerlingen gemeten op basis van een korte toets die materiaal bevatte uit zowel COOL⁵⁻¹⁸ als Inventaar 3VO. Vanwege de overlap is het mogelijk om de toetsscores die de leerlingen behalen op de “Limburg-versie” van de leestoets te vertalen naar de meetschaal die in COOL⁵⁻¹⁸ gebruikt wordt. Andersom geldt precies hetzelfde. De toetsscores zijn ook te vertalen naar de meetschalen van Inventaar 3VO.

De “Limburg-versie” bevatte in totaal 19 items, waarvan er 10 uit COOL⁵⁻¹⁸ kwamen. De betreffende items zijn op papier afgenomen. De overige 9 items kwamen uit Inventaar 3VO. Omdat Inventaar 3VO gebruikmaakt van digitale toetsen zijn deze items via de computer afgenomen. De volgorde van afname is gehusseld. Dit betekent dat de ene groep leerlingen tijdens de afname eerst het papieren gedeelte maakte en daarna doorging met het digitale gedeelte (versie 11 in het databestand), terwijl de andere groep leerlingen in omgekeerde volgorde te werk ging (versie 12 in het databestand). De “Limburg-versie” is gebruikt in alle onderwijstypen. De leerlingen in de basisberoepsgerichte leerweg van het VMBO kregen dus dezelfde items voorgelegd als leerlingen in het HAVO en het VWO. Omdat vooraf niet geheel duidelijk was hoe de items zouden functioneren in de verschillende onderwijstypen, en of de afnameconditie (papier of digitaal)

effect zou hebben op de prestaties van leerlingen, zijn voorafgaand aan de equivalering enkele verkennende analyses verricht.

Eerst is per onderwijstype een toets- en itemanalyse uitgevoerd. Tabel 4.1 bevat de belangrijkste resultaten. De leerlingen die niet alle items gemaakt hadden, zijn buiten de analyse gelaten. Binnen elk van de onderwijstypen staat in de eerste kolom steeds het aantal leerlingen. In de tweede en derde kolom staat respectievelijk de p -waarde (i.e., proportie leerlingen met een correct antwoord) en de r_{it} -waarde. Een hoge r_{it} -waarde wijst erop dat veel leerlingen met een hoge toetsscore het betreffende item correct beantwoorden en dat veel leerlingen met een lage toetsscore het betreffende item fout beantwoorden. Waarden boven de .20 worden in de regel als voldoende aangemerkt.

In Tabel 4.1 is te zien dat de p -waarde bij de meeste items acceptabel is, zij het dat de items voor de leerlingen in de basis- en kaderberoepsgerichte leerweg van het VMBO soms aan de moeilijke kant waren (p -waarde $< .40$) en voor de leerlingen in het VWO soms aan de gemakkelijke kant (p -waarde $> .90$). De r_{it} -waarden liggen veelal ruim boven de .20 en zijn daarmee eveneens acceptabel te noemen. Er is echter één uitzondering. Bij item 15 wist bijna geen enkele leerling het juiste antwoord te geven en bovendien ging een correct antwoord bij dit item vaak niet samen met een hoge toetsscore. Omdat item 15 overduidelijk niet goed functioneerde tijdens de toetsafname is dit item in geen enkele vervolganalyse meegenomen.

Tabel 4.1

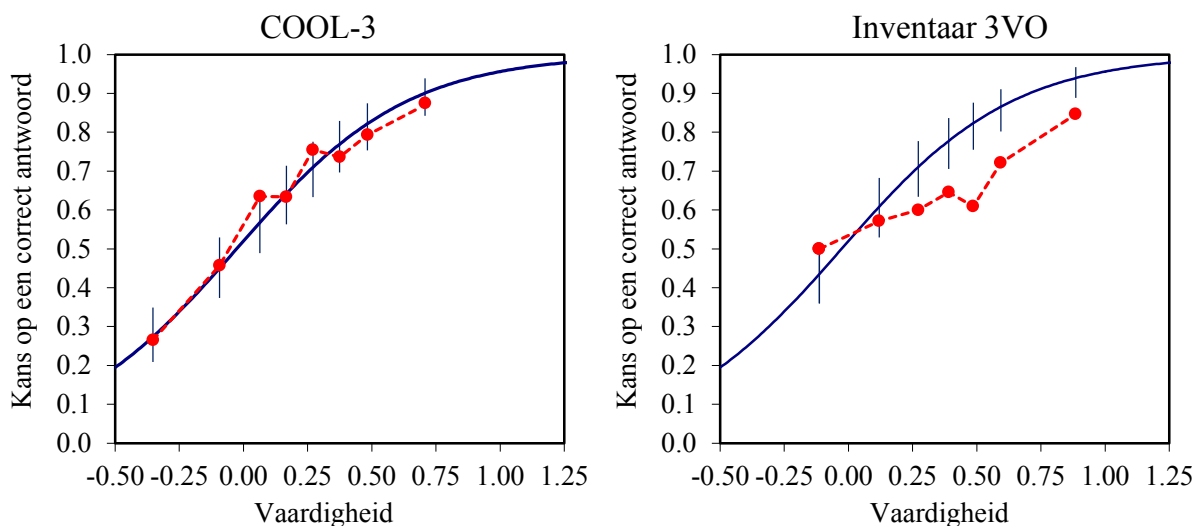
Toets- en itemanalyse begrippend leestoets versies 11 en 12

Item	BBL + KBL			GL + TL			HAVO			VWO		
	n	p	r_{it}	n	p	r_{it}	n	p	r_{it}	n	p	r_{it}
1	126	.61	.25	194	.71	.22	533	.77	.33	371	.84	.13
2	126	.37	.29	194	.70	.33	533	.80	.38	371	.92	.23
3	126	.52	.33	194	.58	.44	533	.74	.42	371	.81	.37
4	126	.68	.43	194	.80	.39	533	.89	.44	371	.96	.45
5	126	.60	.23	194	.73	.20	533	.72	.31	371	.72	.27
6	126	.62	.44	194	.72	.40	533	.84	.32	371	.89	.28
7	126	.71	.42	194	.86	.48	533	.90	.40	371	.97	.29
8	126	.44	.17	194	.53	.25	533	.74	.38	371	.89	.44
9	126	.30	.36	194	.56	.46	533	.63	.24	371	.83	.29
10	126	.37	.20	194	.54	.32	533	.69	.26	371	.89	.41
11	126	.42	.22	194	.55	.33	533	.61	.34	371	.82	.30
12	126	.35	.33	194	.49	.37	533	.67	.41	371	.90	.46
13	126	.61	.58	194	.83	.58	533	.84	.59	371	.96	.59
14	126	.65	.61	194	.76	.51	533	.83	.56	371	.94	.21
15	126	.09	.00	194	.16	-.18	533	.11	.03	371	.08	.12
16	126	.60	.62	194	.80	.52	533	.83	.57	371	.97	.58
17	126	.60	.65	194	.74	.58	533	.79	.54	371	.91	.22
18	126	.39	.31	194	.52	.38	533	.64	.44	371	.88	.20
19	126	.47	.40	194	.61	.32	533	.65	.30	371	.76	.23

In een vervolgstap zijn toetsversies 11 en 12 geanalyseerd met behulp van het OPLM. Deze analyse had als doel om vast te stellen of (a) de papieren en digitale items ondergebracht kunnen worden in één meetschaal en (b) de afnameconditie effect heeft op de prestaties van leerlingen. De prestaties van de leerlingen bleken adequaat beschreven te kunnen worden door het OPLM. Ten eerste bleek de verdeling van p -waarden

(overschrijdingskansen) voor de S -toetsen bij elke categorie redelijk uniform verdeeld te zijn. Ten tweede was de verhouding tussen de R/c -bijdrage en het aantal vrijheidsgraden acceptabel, $R/c = 176$, $df = 118$, $p < .01$. Dit betekent dat de papieren en digitale items hoogstwaarschijnlijk een beroep doen op dezelfde onderliggende vaardigheid en om deze reden inderdaad ondergebracht kunnen worden in één meetschaal. Zodra de papieren en digitale items op één meetschaal zijn geplaatst, wordt het ook mogelijk om de twee afnamecondities met elkaar te vergelijken. Door middel van een gepaarde t -toets is nagegaan in hoeverre leerlingen in de ene conditie een andere vaardigheid laten zien dan in de andere conditie. Als we de prestaties van de leerlingen met behulp van het OPLM corrigeren voor de moeilijkheidsgraad van de items, blijken leerlingen in de papieren ($M = 0.400$, $SD = 0.408$) afnameconditie niet significant hoger of lager te scoren dan in de digitale ($M = 0.389$, $SD = 0.389$) afnameconditie; $t(1327) = 0.896$, $p = 0.370$. Dit resultaat bevestigt nog eens dat het digitale gedeelte equivalent is aan het papieren gedeelte en er daarom met één meetschaal gewerkt kan worden.

Na deze verkennende analyses is de meetschaal voor Limburg gerelateerd aan de meetschaal voor COOL-3. Er zijn verschillende manieren om dat te doen. Een eerste mogelijkheid is om de parameters die eerder geschat zijn voor de items van toetsversies 8, 9 en 10 (zie paragraaf 4.4) te fixeren en de items van toetsversies 11 en 12 er als het ware “bij te schalen”. Deze aanpak bleek niet optimaal te werken. Sommige item karakteristieke curven bleken namelijk niet te passen bij het antwoordgedrag van de Inventaar 3VO leerlingen. Figuur 4.6 geeft een voorbeeld. De blauwe curve representeert het antwoordgedrag zoals we dat zouden verwachten op basis van het OPLM. De rode lijn representeert het gedrag zoals we dat hebben waargenomen tijdens de dataverzameling. Links is te zien dat het OPLM een goede samenvatting geeft van het antwoordgedrag van de COOL-3 leerlingen. De rode lijn ligt immers dicht bij de blauwe lijn. Rechts is te zien dat deze conclusie niet geldt voor de Inventaar 3VO leerlingen. In Inventaar 3VO hebben leerlingen aanzienlijk zwakker gepresteerd dan we op grond van het OPLM mogen verwachten. Als we dit beeld zouden negeren bij het schatten van de vaardigheid van leerlingen bestaat het risico op onder- of overschatting. Daarom is gekozen voor een alternatief.



Figuur 4.6. Antwoordgedrag van leerlingen in COOL-3 en Inventaar 3VO

Om in de equivalering enigszins rekening te kunnen houden met het “afwijkende” antwoordgedrag van de Inventaar 3VO leerlingen is gebruikgemaakt van de zogeheten *proficiency transformation* procedure (Muraki, Hombo, & Lee, 2000; Lee & Ban, 2010). In deze procedure gaan we ervan uit dat we een set items hebben die in twee meetschalen voorkomt. Voor een leerling die de betreffende items maakt, mag het

uiteraard niet uitmaken met welke meetschaal zijn vaardigheid geschat wordt, maar omdat de vaardigheidsschattingen op verschillende meetschalen liggen, zijn ze niet direct vergelijkbaar. Wat we wel direct kunnen vergelijken is de zogeheten z -waarde. De z -waarde geeft de positie van een leerling in de populatie van leerlingen aan. Deze waarde moet onafhankelijk zijn van de meetschaal. Sterker nog: de z -waarde in de ene meetschaal moet exact hetzelfde zijn als de z -waarde in de andere meetschaal. Met dit gegeven is eenvoudig de relatie tussen twee (deels overlappende) meetschalen vast te stellen. Stel dat we de vaardigheid in de Inventaar 3VO meetschaal aanduiden met symbool X . De z -waarde voor deze meetschaal wordt dan gegeven door:

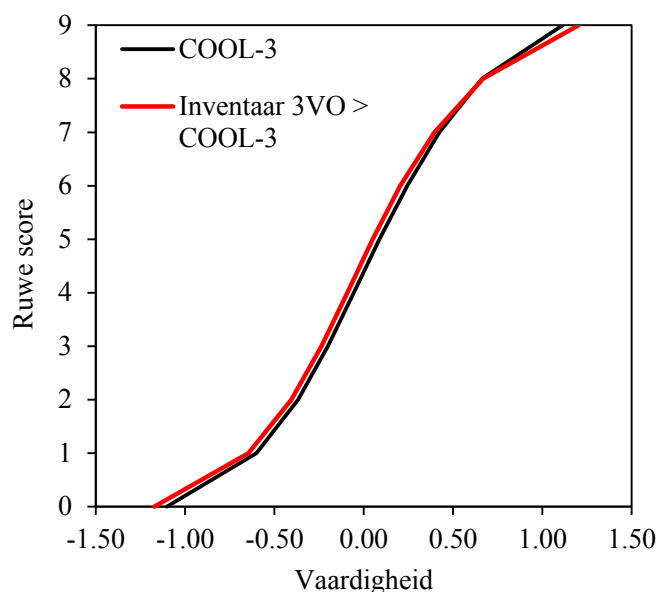
$$\frac{(X - \mu_X)}{\sigma_X},$$

waarbij μ_X het gemiddelde en σ_X de spreiding van de vaardigheid weergeven. Als voor de COOL-3 meetschaal vervolgens het symbool Y gebruikt wordt én aangenomen wordt dat de z -waarden voor beide meetschalen gelijk zijn, dan moet gelden dat:

$$\frac{(X - \mu_X)}{\sigma_X} \stackrel{\text{def}}{=} \frac{(Y - \mu_Y)}{\sigma_Y}, \text{ oftewel } Y = \mu_Y + \frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

Op deze manier is er dus een directe omzetting van de Inventaar 3VO meetschaal naar de eerder geconstrueerde COOL-3 meetschaal mogelijk.

Figuur 4.7 laat zien hoe de procedure uitpakt bij de koppeling van Inventaar 3VO aan COOL-3. De zwarte lijn representeert de relatie tussen de vaardigheidsscores en de (ongewogen) ruwe scores zoals die geldt in COOL-3. De rode lijn representeert dezelfde relatie, maar dan voor Inventaar 3VO ná toepassing van de hiervoor beschreven transformatie. Bij het construeren van de lijnen zijn vanzelfsprekend alleen de overlappende items meegenomen. De rode lijn zou exact over de zwarte lijn heen liggen als de Inventaar 3VO leerlingen hetzelfde antwoordgedrag vertonen als de COOL-3 leerlingen. De rode lijn kan verder van de zwarte lijn af komen te liggen naarmate de verschillen in antwoordgedrag groter zijn. Zoals al duidelijk werd in Figuur 4.6 verschilt het antwoordgedrag in Inventaar 3VO soms van die in COOL-3. Desalniettemin vertoont de rode lijn grote gelijkenis met de zwarte lijn en lijkt de *proficiencytransformation* procedure een verdedigbare koppeling tussen COOL-3 en Inventaar 3VO op te leveren. Op basis van deze transformatie zijn dan ook de (geëquivalenteerde) scores van de leerlingen bepaald.



Figuur 4.7. *Vergelijking van Inventaar 3VO en COOL-3 op basis van de toets karakteristieke curve*

Net als eerder bij toetsversies 8, 9 en 10 bevat het databestand voor toetsversies 11 en 12 de ruwe scores (**BGL_RSCO**, max = 18), de gewogen scores (**BGL_WSCO**, max. = 62), de vaardigheidsscores (**BGL_THET**) en de bankscores (**BGL_BANK3**, max. = 100). De ruwe en gewogen scores hebben betrekking op toetsversies 11 en 12 en kunnen niet vergeleken worden met de ruwe en gewogen scores van de andere toetsversies. De vaardigheidsscores zijn in eerste instantie geschat op de Inventaar 3VO meetschaal. Vervolgens zijn de vaardigheidsscores via de eerder beschreven lineaire transformatie gerelateerd aan de COOL-3 meetschaal, zodat het mogelijk wordt om vergelijkingen te maken tussen Inventaar 3VO en COOL-3. De bankscores zijn berekend op basis van de lineair getransformeerde vaardigheidsscores en de itemparameters zoals die zijn geschat op basis van de COOL-3 data. De 9 items van Inventaar 3VO zijn niet meegenomen bij het berekenen van de bankscores.

Hoewel Figuur 4.7 erop wijst dat de koppeling die gemaakt is tussen Inventaar 3VO en COOL-3 adequaat is, moeten de bank- en vaardigheidsscores zoals die berekend zijn voor toetsversies 11 en 12 wel met de nodige voorzichtigheid gebruikt worden:

1. De toets die in Limburg gebruikt is, was tamelijk kort en de overlap met COOL-3 was beperkt. Uiteindelijk is de koppeling tussen Inventaar 3VO en COOL-3 gebaseerd op 9 items. Een groter anker leidt in de regel tot een stabielere (of nauwkeurigere) equivalering, maar dat was in het onderhavige onderzoek om praktische redenen niet te realiseren.
2. Het antwoordgedrag van de leerlingen die deelnamen aan Inventaar 3VO week soms af van het antwoordgedrag dat werd waargenomen in COOL-3. Door voor Inventaar 3VO een aparte meetschaal te construeren en deze pas in een vervolgstap via de zogeheten *proficiencytransformation* procedure te relateren aan de meetschaal van COOL-3 konden we het risico op bias in de vaardigheidsscores beperken. De lineaire transformatie doet immers geen afbreuk aan de kwaliteit van de vaardigheidsscores. Dit geldt echter alleen voor analyses *binnen* Inventaar 3VO. Als een vergelijking gemaakt wordt *tussen* Inventaar 3VO en COOL-3 speelt de lineaire transformatie wel een rol. De transformatie lijkt verdedigbaar, maar uiteindelijk is het vanwege de waargenomen verschillen in antwoordgedrag niet geheel zeker of de vergelijking tussen Inventaar 3VO en COOL-3 ook daadwerkelijk adequaat te maken is.

3. Bij het berekenen van de bankscores is gebruikgemaakt van de lineair getransformeerde vaardigheidsscores en de itemparameters van COOL-3. De vaardigheidsscores zijn gebaseerd op 18 items. De betrouwbaarheid is in het licht van de toetslengte goed te noemen (.79) en ook uit de toetsinformatiefunctie blijkt dat toetsversies 11 en 12 goed aansloten bij het vaardigheidsniveau van de leerlingen die de toetsen maakten. Alleen voor de (betere) leerlingen in het HAVO en VWO was de toets aan de gemakkelijke kant en zijn de vaardigheidsscores behept met een grotere meetfout. Dit heeft uiteraard ook zijn weerslag op de bankscores, maar omdat analyses doorgaans op groepsniveau plaatsvinden is de iets grotere meetfout aan de bovenkant van de meetschaal geen groot probleem. Wel blijft staan dat de bankscores gebaseerd zijn op de itemparameters van COOL-3 en dat daarmee verondersteld is dat deze geldig zijn voor Inventaar 3VO. Zoals al eerder is aangegeven, is deze veronderstelling alleen te controleren voor de 9 ankeritems. Of deze veronderstelling ook te verdedigen is voor de overige 66 items is niet te onderzoeken.

4.8 Koppeling met het Cito volgsysteem VO

Bij de start van COOL⁵⁻¹⁸ in schooljaar 2007/2008 werd in het voortgezet onderwijs slechts beperkt gebruikgemaakt van leerlingvolgsystemen. De afgelopen jaren is het gebruik hiervan sterk toegenomen. Om deze reden zijn bij de afgelopen meting, bij wijze van proef, op een beperkt aantal scholen, de toetsen van het Cito Volgsysteem VO als alternatief gebruikt voor de COOL⁵⁻¹⁸-toetsen. In essentie kwam het erop neer dat de betreffende scholen via een overeenkomst toestemming gaven om de toetsgegevens die zij in schooljaar 2014/2015 verzamelden met behulp van het Cito Volgsysteem VO (geanonimiseerd) op te nemen in het COOL⁵⁻¹⁸-databestand. In totaal hebben 7 scholen voor het voortgezet onderwijs toestemming gegeven voor het gebruik van hun leerlingvolgsysteemgegevens.

Het Cito Volgsysteem VO bevat voor het derde leerjaar drie verschillende leestoeten. De leestoetsen zijn bedoeld voor verschillende onderwijstypen. De eerste toets bevat 45 items en is bedoeld voor leerlingen in de basisberoepsgerichte leerweg van het VMBO. In het databestand is aan deze toets nummer 13 toegekend. De tweede toets (**BGL_VERS** = 14) bevat 51 items en is bedoeld voor leerlingen in de kaderberoepsgerichte, en gemengde en theoretische leerweg van het VMBO. De derde toets wordt afgenomen bij HAVO en VWO leerlingen (**BGL_VERS** = 15) en bevat 54 items. De items in de leestoetsen liggen niet op dezelfde meetschaal als de items van COOL⁵⁻¹⁸. Dit betekent dat we de prestaties die leerlingen behalen op het Cito Volgsysteem VO niet direct kunnen vergelijken met de prestaties van de leerlingen die de COOL⁵⁻¹⁸-toetsen hebben gemaakt. Door toepassing van een equivaleringsprocedure zijn er echter wel mogelijkheden om het Cito Volgsysteem VO te relateren aan COOL⁵⁻¹⁸. Vaak vindt een equivalering plaats op basis van een set gemeenschappelijke items (zie ook paragraaf 4.2), maar deze aanpak was niet toe te passen, omdat het Cito Volgsysteem VO geen (of hooguit enkele) items bevat die ook in COOL⁵⁻¹⁸ gebruikt worden. Een alternatief is om de gemeenschappelijkheid te zoeken in de leerlingen (zie Eggen, 1993), maar ook deze aanpak bleek weinig kansrijk. Bij toepassing van deze aanpak zou de relatie namelijk gelegd worden via een groep leerlingen die zowel de COOL⁵⁻¹⁸-toetsen als de volgsysteemtoetsen maakt. Dit kwam onvoldoende voor. Daarom is van deze aanpak afgezien.

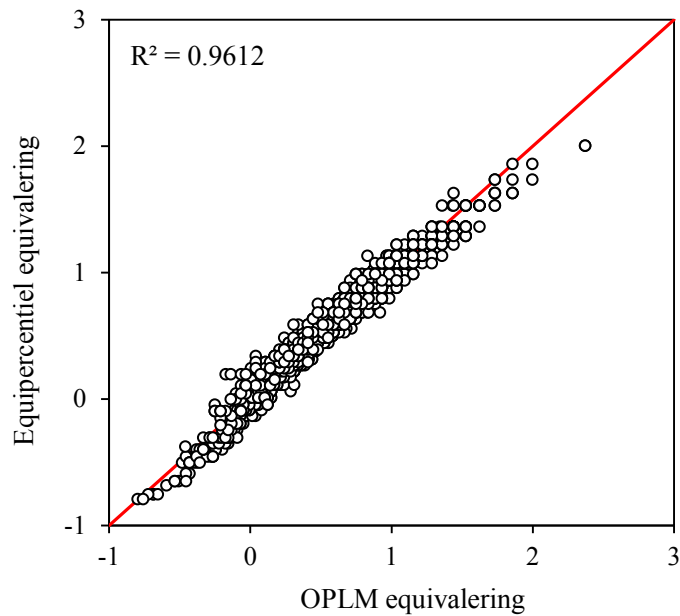
Indien items (of toetsen) niet op dezelfde meetschaal geplaatst kunnen worden via gemeenschappelijke items of leerlingen, is een equivalering alleen mogelijk onder een aantal aannames. In het onderhavige geval is verondersteld dat de leestoetsen van COOL⁵⁻¹⁸ en het Cito Volgsysteem VO gemaakt zijn door statistisch equivalente groepen. Er is, met andere woorden, aangenomen dat de leerlingen die de volgsysteemtoetsen maakten uit dezelfde onderliggende vaardigheidsverdeling afkomstig zijn als de leerlingen die de COOL⁵⁻¹⁸-toetsen maakten. Een dergelijke aanname lijkt zonder meer verdedigbaar te

zijn, omdat de inhoud van de volgsysteemtoetsen grote gelijkenis vertoont met de inhoud van de COOL⁵⁻¹⁸-toetsen. Bovendien zaten de leerlingen ten tijde van de afname allemaal in de tweede helft van het derde leerjaar van het VO. Onder aanname van statistisch equivalente groepen is het mogelijk om het OPLM te schatten met behulp van *marginal maximum likelihood* (MML). We houden bij het schatten van de itemparameters dan rekening met de groep leerlingen die een toetsboekje maakt en doen daarnaast een veronderstelling over de verdeling van de vaardigheid in de populatie. Wij hebben aangenomen dat de vaardigheid in de populatie normaal verdeeld is.

In totaal hebben de 7 scholen van 1567 leerlingen toetsgegevens ingestuurd. Sommige leerlingen bleken alleen deelgenomen te hebben aan de onderdelen wiskunde, Engels of taalverzorging en geen leestoets te hebben gemaakt. Uiteindelijk konden we per toetsversie beschikken over de volgende hoeveelheid gegevens: (**BGL_VERS 13**) 17 leerlingen, (**BGL_VERS 14**) 475 leerlingen, en (**BGL_VERS 15**) 910 leerlingen. Het is opmerkelijk dat de toetsversie die bedoeld is voor de basisberoepsgerichte leerweg van het VMBO slechts door 17 leerlingen is gemaakt. De meeste leerlingen in de basisberoepsgerichte leerweg bleken een versie van een hoger niveau gemaakt te hebben. Met 17 waarnemingen is er geen stabiele basis om de toets te relateren aan COOL⁵⁻¹⁸. Er is dan ook besloten om daarvan af te zien. Bij de twee andere toetsversies hadden we wel voldoende gegevens en is het OPLM geschat met behulp van MML. Uit de statistische toetsen bleek dat het OPLM een adequate beschrijving geeft van de prestaties van de leerlingen: $R^2 = .411$, $df = 412$, $p = .51$. Het lijkt dan ook zonder meer mogelijk om de scores die de leerlingen behaald hebben op de leestoetsen van het Cito Volgsysteem VO met behulp van dit model te vertalen naar de meetschaal van COOL⁵⁻¹⁸.

Hoewel de equivalering op basis van het OPLM verdedigbaar lijkt, vermengen we bij toepassing van MML wel twee modellen: het OPLM vertelt iets over de antwoorden gegeven θ en de normale verdeling vertelt hoe de θ 's in de populatie verdeeld zijn. De verstrengeling van beide modellen gebeurt op een zodanig diep niveau dat ze niet uit elkaar te halen zijn. Maken we een fout in de veronderstelling over de normale verdeling (θ is niet normaal verdeeld, dan wel θ is normaal verdeeld, maar onze steekproef is niet aselekt uit de verdeling getrokken), dan heeft dat tot gevolg dat er systematische fouten geïntroduceerd worden in de schattingen van de itemparameters. Uiteraard heeft deze bias ook invloed op de vaardigheidsschattingen en daarmee eveneens op vergelijking tussen het Cito Volgsysteem VO en COOL⁵⁻¹⁸. Om deze reden is de equivalering gecontroleerd op basis van een (klassieke) equipercentielevaivering zonder smoothing (zie bijvoorbeeld Kolen & Brennan, 2004). Bij een dergelijke equivalering veronderstellen we dat het relatieve niveau van leerlingen bij het Cito Volgsysteem VO gelijk is aan relatieve niveau dat zij zouden behalen bij deelname aan COOL⁵⁻¹⁸. Hoewel deze aanname behoorlijk stringent is en de systematiek fundamenteel verschilt van het OPLM mogen we wel verwachten dat de equipercentielevaivering gelijkenis vertoont met de OPLM equivalering.

Bij de equipercentielevaivering is geen onderscheid gemaakt in onderwijstypen. Dit betekent dat het relatieve niveau van leerlingen is bepaald in de totale groep met leerlingen die toetsversie 14 of 15 hebben gemaakt. Figuur 4.8 laat zien hoe de equipercentielevaivering zich verhoudt tot de equivalering met het OPLM. We zien dat er sprake is van een sterke lineaire samenhang, $r = .98$. Voor het bepalen van de geëquivalenteerde scores van de leerlingen maakt het dus weinig uit welke aanpak we kiezen. Wij hebben ervoor gekozen om gebruik te maken van de scores die voortvloeien uit de OPLM equivalering.



Figuur 4.8. *Equipercentielevaivering vergeleken met een OPLM equivalering*

Net als eerder bij toetsversies 8, 9, 10, 11 en 12 bevat het databestand voor toetsversies 14 en 15 de ruwe scores (**BGL_RSCO**, max. versie 14 = 51 / max. versie 15 = 54), de gewogen scores (**BGL_WSCO**, max. versie 14 = 62 / max. versie 15 = 95), de vaardigheidsscores (**BGL_THET**) en de bankscores (**BGL_BANK3**, max. = 100). De ruwe en gewogen scores hebben betrekking op toetsversies 14 en 15 en kunnen niet vergeleken worden met de ruwe en gewogen scores van de andere toetsversies. Op basis van de bank- en vaardigheidsscores kan die vergelijking wel gemaakt worden. Het heeft de voorkeur om bij analyses *binnen* een toetsversie gebruik te maken van de ruwe scores. Analyses over toetsversies of onderwijstypen heen vinden bij voorkeur plaats op basis van de bankscores die in het databestand zijn opgenomen.

5 TAALVERZORGING

5.1 Materiaal

Bij de meting in 2014 is taalverzorging op dezelfde manier getoetst als bij de eerdere meting in 2011. Dit betekent dat de volgende categorieën aan de orde kwamen tijdens de toetsafname: *grammatica*, *werkwoordspelling* en *niet-werkwoordspelling*. Bij de meeste items konden leerlingen kiezen uit vier alternatieven. De opdracht was vaak: “In welke zin is het dikgedrukte woord fout gespeld? Streep de letter van deze zin aan op het antwoordblad”. Bij werkwoordspelling stond de d/t-regel bij werkwoorden in de tegenwoordige en verleden tijd centraal. Daarnaast werd gevraagd naar de spelling van het voltooid deelwoord en de infinitief. Bij niet-werkwoordspelling ging het om zowel orthografische als morfologische en logografische spellingproblemen. Denk hierbij bijvoorbeeld aan woorden met medeklinkers die niet worden uitgesproken, bijvoeglijke naamwoorden met of zonder –n– aan het eind, de letters g/ch en ei/ij, uitheemse woorden of woorden met een trema. Hieronder volgt voor beide vormen van spelling een voorbeelditem:

Werkwoorden **In welke zin is het vetgedrukte woord fout geschreven?**

- A De kok **snijd** zich in zijn vinger.
- B Mijn oma heeft even **gerust**.
- C De jongen **jankte** van woede.
- D **Juichend** kwam de schaatser over de finish.

Niet-werkwoorden **In welke zin is het vetgedrukte woord fout geschreven?**

- A Bij het inademen van de rook krijg je een **benauwd** gevoel.
- B Hij houdt niet zo van vis en al zeker niet van **kabeljauw**.
- C Ik vind het hier beneden in de kelder **ijskoud**.
- D De thee was ondertussen **louw** geworden.

Bij de categorie grammatica werd gevraagd naar de vorming van voornaamwoorden en onregelmatige werkwoorden. Daarnaast werd getoetst of leerlingen de grammaticale begrippen die in het referentiekader genoemd worden beheersen. Het gaat hierbij om lidwoorden, bijvoeglijke naamwoorden, voornaamwoorden en persoonsvormen. Hieronder volgt een voorbeelditem:

Grammatica Overal waar **we** gaan **of** staan komen we camera’s tegen.

Er is **zelfs** cameratoezicht in **het** openbaar vervoer.

Welk vetgedrukt woord is een lidwoord?

- A **we**
- B **of**
- C **zelfs**
- D **het**

In Zijssling et al. (2013) is een uitgebreidere beschrijving van de inhoud van de toetsen voor taalverzorging te vinden.

5.2 Onderzoeksdesign

Tijdens de meting in 2014 zijn in totaal 80 items afgenomen. De items zijn afgenomen volgens een structureel onvolledig design. Figuur 5.1 geeft het design zoals dat gebruikt is tijdens de afname. Eerst zijn de items verdeeld in vier modules met 20 items (A tot en met D). Vervolgens zijn per onderwijstype twee modules geselecteerd die qua inhoud en moeilijkheidsgraad geschikt leken voor afname. VMBO leerlingen kregen afhankelijk van de leerweg modules A, B, of C voorgelegd en HAVO en VWO leerlingen kregen modules C en D voorgelegd. In het design is er voor gezorgd dat er zogenaamde ankermodules waren. Dat betekent dat een bepaalde groep leerlingen steeds een aantal items gemeenschappelijk had met een andere groep leerlingen. In het databestand geeft de variabele **TVZ_VERS** aan welke combinatie van modules aan de leerling is voorgelegd.

Versie	Onderwijstype	Module			
		A	B	C	D
7	BBL				
8	KBL/GL/TL				
9	HAVO/VWO				

Figuur 5.1. *Afnamedesign taalverzorging meting 3*

Zoals in paragraaf 5.1 al is opgemerkt is taalverzorging bij de tweede meting van COOL⁵⁻¹⁸ in 2011 op exact dezelfde manier getoetst. De scores die door de leerlingen behaald zijn bij de twee metingen kunnen echter *niet* met elkaar vergeleken worden, omdat er in COOL-3 voor gekozen is om de itemantwoorden van de leerlingen opnieuw te analyseren. Het gevolg van deze keuze is dat de gewogen scores en de vaardigheidsscores voor COOL-3 voortvloeien uit een ander meetmodel dan in COOL-2 en er zodoende geen basis is voor het maken van een vergelijking tussen de twee metingen. Vanzelfsprekend zijn er wel mogelijkheden om de metingen aan elkaar te relateren. Deze koppeling wordt echter niet in dit rapport, maar in een aanvullend rapport beschreven (Keuning, Keizer-Mittelhaeuser, & Timmermans, in voorbereiding). De scores die in dit rapport beschreven worden, kunnen dus alleen gebruikt worden om analyses te verrichten *binnen* de derde meting van COOL⁵⁻¹⁸ in VO-3. Vergelijkingen met de tweede meting van COOL⁵⁻¹⁸ kunnen *niet* gemaakt worden op basis van dit rapport en het bijbehorende databestand. Om dit nog eens extra te benadrukken worden in het databestand de versie nummers 7, 8 en 9 gebruikt in plaats van de eerdere versie nummers 4, 5 en 6.

5.3 Meetmodel

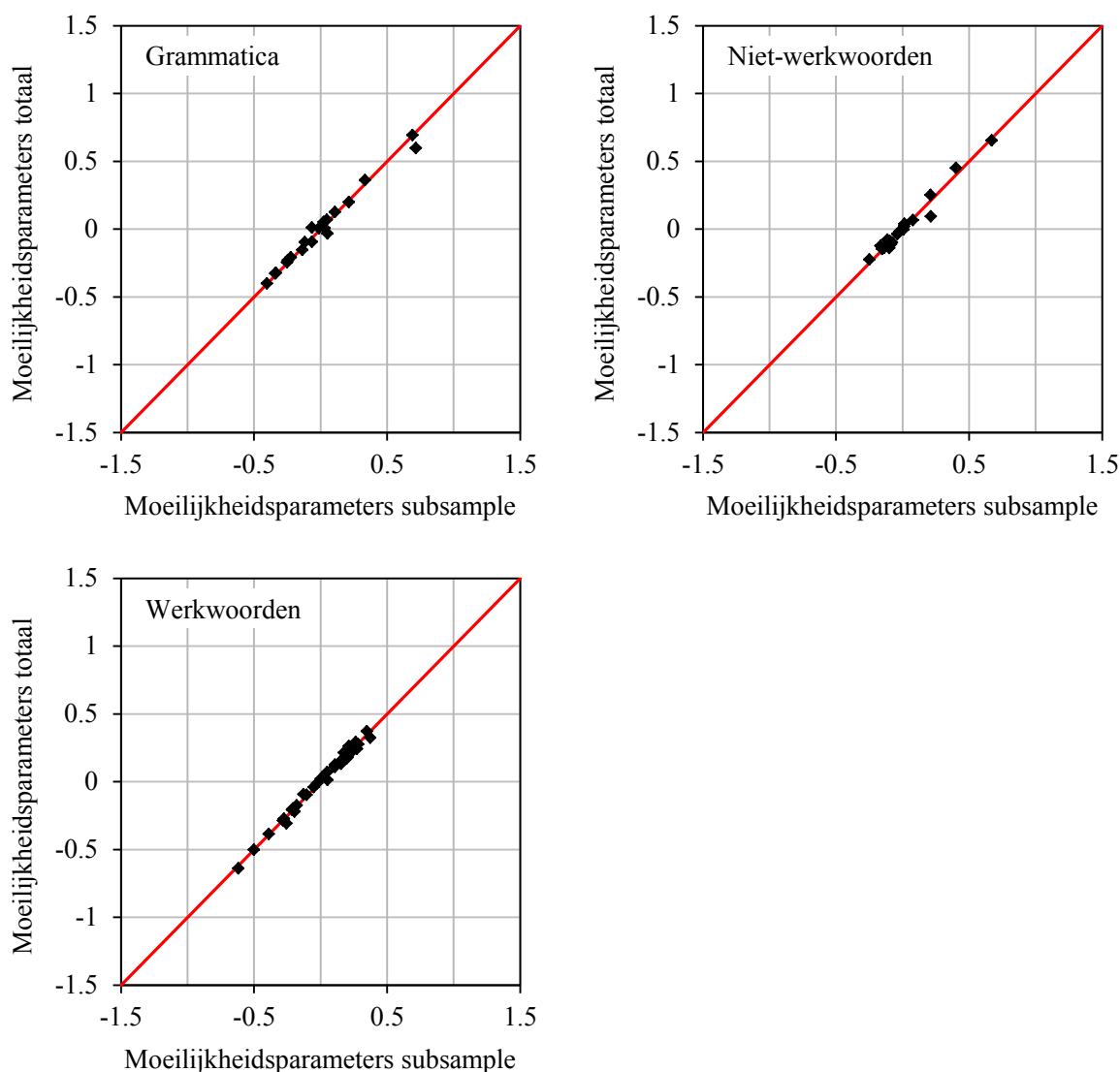
In grote lijnen is het onderdeel taalverzorging op dezelfde wijze geanalyseerd als het onderdeel begrijpend lezen. Dit betekent dat gebruikgemaakt is van het OPLM voor dichtome items. Paragraaf 4.3 gaat in op de belangrijkste kenmerken en eigenschappen van het OPLM. De analyse voor taalverzorging verschilt echter op één belangrijk punt van de analyse voor begrijpend lezen. Waar de items voor begrijpend lezen in 1x geanalyseerd zijn, heeft de analyse bij taalverzorging per categorie plaatsgevonden. Het OPLM is dus driemaal geschat: 1x voor grammatica, 1x voor werkwoordspelling en 1x voor niet-werkwoordspelling. De geëquivalenteerde scores voor de leerlingen per categorie zijn in een vervolgstap samengevoegd tot een zogenaamde *compositescore* (zie paragraaf 5.6).

Bij taalverzorging is gekozen voor een analyse (of equivalering) per categorie, omdat de aanname van unidimensionaliteit niet goed houdbaar was. In geval de multidimensionale structuur bij taalverzorging genegeerd zou zijn, en er dus toch één ‘overall’ analyse met het OPLM zou hebben plaatsgevonden, zouden de vaardigheidsscores en de daaruit voortvloeiende bankscores voor de leerlingen mogelijk een onrealistische weergave zijn geweest van de werkelijkheid. Dit probleem wordt ondervangen door het OPLM per categorie te schatten.

5.4 Kalibratie

Net zoals bij begrijpend lezen varieerde het aantal waarnemingen bij taalverzorging als gevolg van het afnamedesign sterk. De items in module A zijn bijvoorbeeld door 409 leerlingen gemaakt, terwijl de items in module C door 6986 leerlingen zijn gemaakt. Hoewel het OPLM zonder problemen geschat kan worden met een dergelijke verdeling van waarnemingen, zal het beoordelen van de fit van het OPLM problematisch zijn (zie ook paragraaf 4.4). Om problemen bij het beoordelen van de passing van het model te vermijden, is de analyse in twee stappen uitgevoerd. In de eerste stap is het OPLM per categorie geschat op basis van alle beschikbare waarnemingen. In een vervolgstap is het OPLM per categorie geschat op basis van een random *sample* uit de totale dataset waarin het aantal waarnemingen per item varieerde van minimaal 400 tot maximaal 800. Bij de beoordeling van de passing van het OPLM is in eerste instantie uitgegaan van de resultaten zoals die verkregen werden op basis van de *subsample*. Daarna is de samenhang tussen de schattingen op basis van de totale dataset en de schattingen op basis van de *subsample* in kaart gebracht. De uiteindelijke kalibratie is gebaseerd op de dataset met alle waarnemingen.

De kalibratie op basis van de *subsample* liet zien dat de prestaties van de leerlingen op de drie categorieën binnen taalverzorging adequaat beschreven kunnen worden door het OPLM. Ten eerste bleek de verdeling van *p*-waarden (overschrijdingskansen) voor de *S*-toetsen bij elke categorie redelijk uniform verdeeld te zijn. Ten tweede was de verhouding tussen de *Rlc*-bijdrage en het aantal vrijheidsgraden acceptabel, (grammatica) $Rlc = 73$, $df = 78$, $p = .64$; (niet-werkwoorden) $Rlc = 134$, $df = 88$, $p < .01$; (werkwoorden) $Rlc = 210$, $df = 149$, $p < .00$. De kalibraties op basis van de *subsample*, ten slotte, vertoonde grote gelijkenis met de kalibraties op basis van de dataset met alle waarnemingen. Figuur 5.2 laat duidelijk zien dat bij elke categorie geldt dat de relatie tussen de twee sets met moeilijkheidsparameters lineair is.



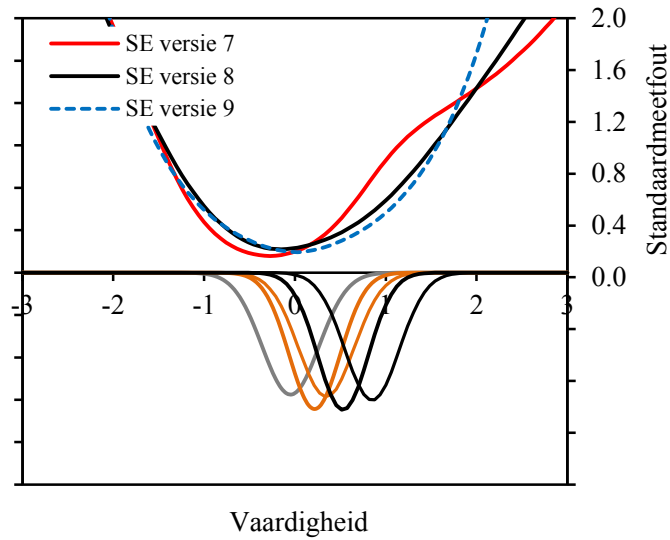
Figuur 5.2. Relatie tussen moeilijkheidsparameters in twee verschillende kalibraties

Bij twee items binnen de categorie *niet-werkwoorden* bleken de prestaties van leerlingen niet goed beschreven te kunnen worden door het OPLM (nummer 13 en 31 en in versie 8; nummer 30 en 32 in versie 9). Ditzelfde probleem speelde bij vier items binnen de categorie *grammatica* (nummer 23 en 34 in versie 7; nummer 33, 34 en 38 in versie 8; nummer 37 en 40 in versie 9). De betreffende items zijn uit de meetschaal gelaten.

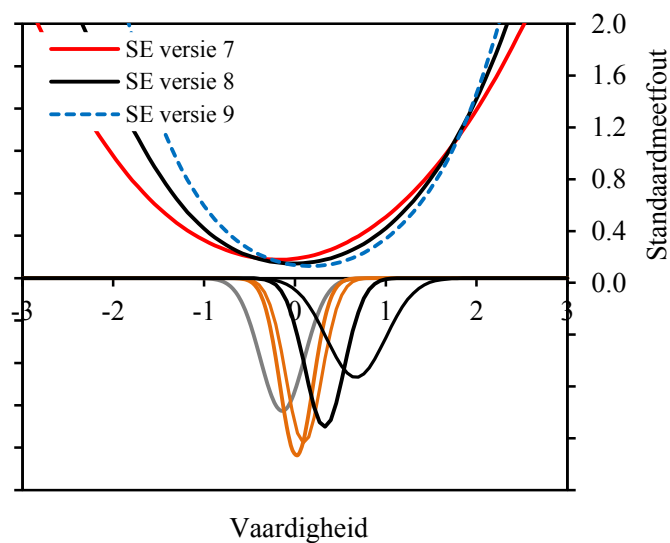
5.5 Meetnauwkeurigheid

Figuren 5.3a, 5.3b en 5.3c geven voor elke categorie en voor elke toetsversie de standaardmeetfout. Onder de x-as staan de kansdichtheidsfuncties (de ongewogen vaardigheidsverdelingen zoals geschat via *marginal maximum likelihood* met gefixeerde itemparameters) van de groepen leerlingen die de verschillende toetsversies gemaakt hebben. In verband met de leesbaarheid van Figuur 5.3 is niet aangegeven op welk onderwijstype de kansdichtheidsfuncties betrekking hebben. Dit is ook niet per se nodig, omdat de gemiddelde vaardigheid – zoals verwacht mag worden – consequent oploopt over de verschillende onderwijstypen. Links op de vaardigheidsschaal vinden we dus de kansdichtheidsfunctie voor de BBL-leerlingen en rechts op de vaardigheidsschaal voor de VWO-leerlingen. We zien dat toetsversie 7 het

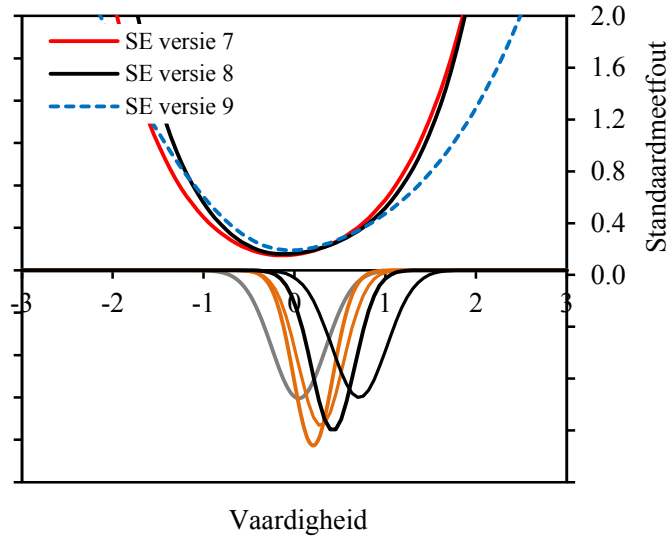
nauwkeurigst meet in de ‘lagere’ onderwijstypen en toetsversie 9 in de ‘hogere’ onderwijstypen. Tegelijkertijd zien we dat de toetsen wat te gemakkelijk zijn geweest voor veel leerlingen. Met name bij de betere leerlingen in het VWO is de meetfout aan de grote kant in vergelijking met de meetfout bij andere leerlingen.



Figuur 5.3a. *Standaardmeetfout grammatica*



Figuur 5.3b. *Standaardmeetfout werkwoordspelling*



Figuur 5.3c. *Standaardmeetfout niet-werkwoordspelling*

Omdat er geen criteria beschikbaar zijn om de lokale meetnauwkeurigheid zoals die is weergegeven in Figuur 5.3 te interpreteren, is voor elke toetsversie ook de klassieke betrouwbaarheid bepaald. Eerst is per categorie de *MAcc* coëfficiënt in de totale populatie bepaald onder de veronderstelling dat alle k_0 items binnen een categorie gemaakt zijn door alle leerlingen. Vervolgens is met behulp van de Spearman-Brown-formule een schatting gemaakt van de betrouwbaarheid in geval een toets t uit k_t items bestaat. Tabel 5.2 geeft per categorie de geschatte betrouwbaarheid uitgesplitst naar toetsversie.

Tabel 5.2

Betrouwbaarheid uitgesplitst naar categorie en toetsversie

Categorie	k_0	ρ_0	Versie 4			Versie 5			Versie 6		
			k_t	σ_t	ρ_t	k_t	σ_t	ρ_t	k_t	σ_t	ρ_t
Grammatica	20	.83	10	14.7	.71	9	12.6	.69	10	7.3	.71
Niet-werkwoorden	22	.83	12	17.6	.72	10	13.7	.68	10	10.8	.68
Werkwoorden	32	.90	16	19.4	.81	16	18.1	.81	16	16.7	.81

De leerlingen worden echter niet gescoord per categorie, maar op het totaal. Daarom is in de laatste stap een schatting gemaakt van de betrouwbaarheid voor de *composite score*. Als composite L gegeven wordt door:

$$L = \sum_{i=1}^n w_i X_i,$$

met X_i de score en w_i het gewicht voor categorie i , en we veronderstellen dat de foutenvarianties tussen de categorieën lineair onafhankelijk zijn, dan wordt de composite betrouwbaarheid ρ gegeven door (Feldt & Brennan, 1989; Thissen & Wainer, 2001; Webb, Shavelson, & Haertel, 2007):

$$\rho = \frac{\sum_{i=1}^n w_i^2 \rho_i \sigma_{X_i}^2 + \sum_{i=1}^n \sum_{q(\neq i)=1}^n w_i w_q \rho_{i,q} \sigma_{X_i} \sigma_{X_q}}{\sum_{i=1}^n w_i^2 \sigma_{X_i}^2 + \sum_{i=1}^n \sum_{q(\neq i)=1}^n w_i w_q \rho_{i,q} \sigma_{X_i} \sigma_{X_q}}.$$

Zoals we kunnen zien, hebben we naast de gewichten, de betrouwbaarheid en de standaarddeviaties per categorie ook de correlaties nodig om de betrouwbaarheid te kunnen bepalen. Tabel 5.3 geeft per toetsversie de correlaties tussen de geobserveerde scores van de leerlingen voor de drie categorieën. Gegeven deze correlaties en de informatie uit Tabel 5.2 is de composite betrouwbaarheid voor toetsversie 4 met 38 items gelijk aan .88. De betrouwbaarheid voor toetsversie 5 is gelijk aan .86 (35 items) en de betrouwbaarheid voor toetsversie 6 is gelijk aan .88 (36 items).

Tabel 5.3

Correlaties tussen werkwoorden (WW), niet-werkwoorden (NW) en grammatica (GR)

	Versie 7			Versie 8			Versie 9		
	WW	NW	GR	WW	NW	GR	WW	NW	GR
WW	1			1			1		
NW	.59	1		.53	1		.57	1	
GR	.54	.66	1	.40	.58	1	.54	.64	1

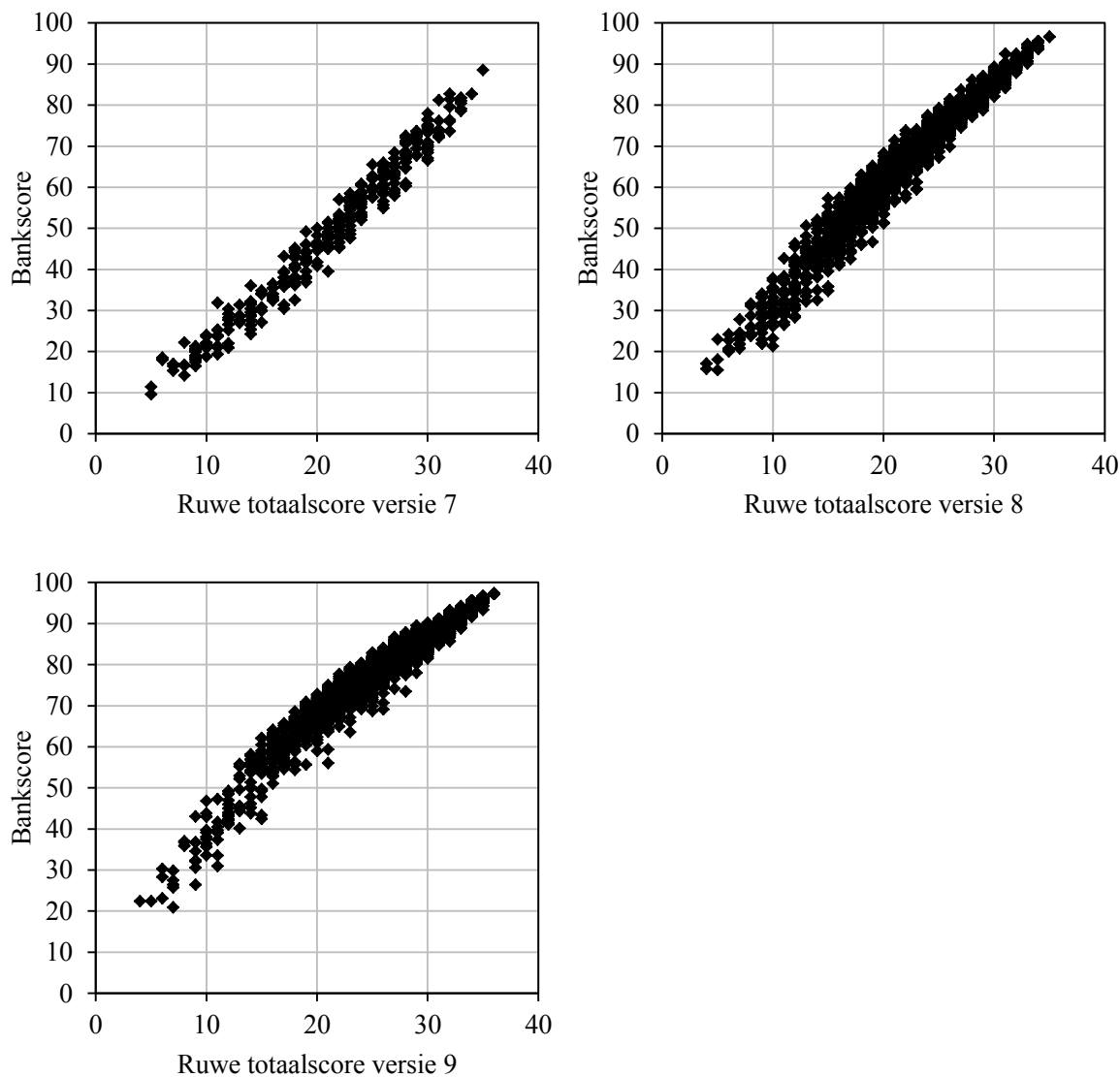
5.6 Toetsresultaten

Het databestand bevat voor het onderdeel Taalverzorging de ruwe totaalscore en een composite score (of bankscore). De betreffende variabelen heten respectievelijk **TVZ_RSCO** en **TVZ_BANK3**. Op basis van de ruwe totaalscores kunnen geen vergelijkingen gemaakt worden over toetsversies heen. Met de bankscore is dat wel mogelijk. Zoals eerder aangegeven is de bankscore ontstaan na een weging van de (deel)scores op de categorieën *werkwoorden*, *niet -werkwoorden* en *grammatica*. Meer specifiek, als het gewicht w voor een bepaalde categorie i met k items gegeven wordt door:

$$w_i = \frac{\sum_{i=1}^n \sum_{j=1}^{k_i} \alpha_j}{n \times \sum_{j=1}^{k_i} \alpha_j}, \quad X_i = \sum_{j=1}^{k_i} \alpha_j p_j(\theta_i) \quad \text{en de score } X_i \text{ door} \quad , \text{ dan is de bankscore } L \text{ gelijk aan:}$$

$$L = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n \sum_{j=1}^{k_i} \alpha_j} \times 100$$

Bij het bepalen van de ruwe totaalscore en de bankscore zijn alleen de geschaalde items meegenomen. Dit betekent dat de maximale ruwe score voor versie 4 gelijk is aan 38, voor versie 5 gelijk aan 35 en voor versie 6 gelijk aan 36. De bankscores liggen tussen 0 en 100. Figuur 5.4 geeft een visuele weergave van de relatie tussen de ruwe totaalscore en de bankscore L . Vanzelfsprekend bestaat er geen één op één relatie tussen de twee scores. Dat komt doordat (a) de weging van de afzonderlijke categorieën en de items verschilde bij de berekening van beide scores, en (b) er bij de berekening van de bankscores een correctie heeft plaatsgevonden voor de moeilijkheidsgraad van de toetsen.



Figuur 5.4. *Visuele weergave van de relatie tussen ruwe totaalscore en de bankscore*

5.7 Koppeling met het Cito volgsysteem VO

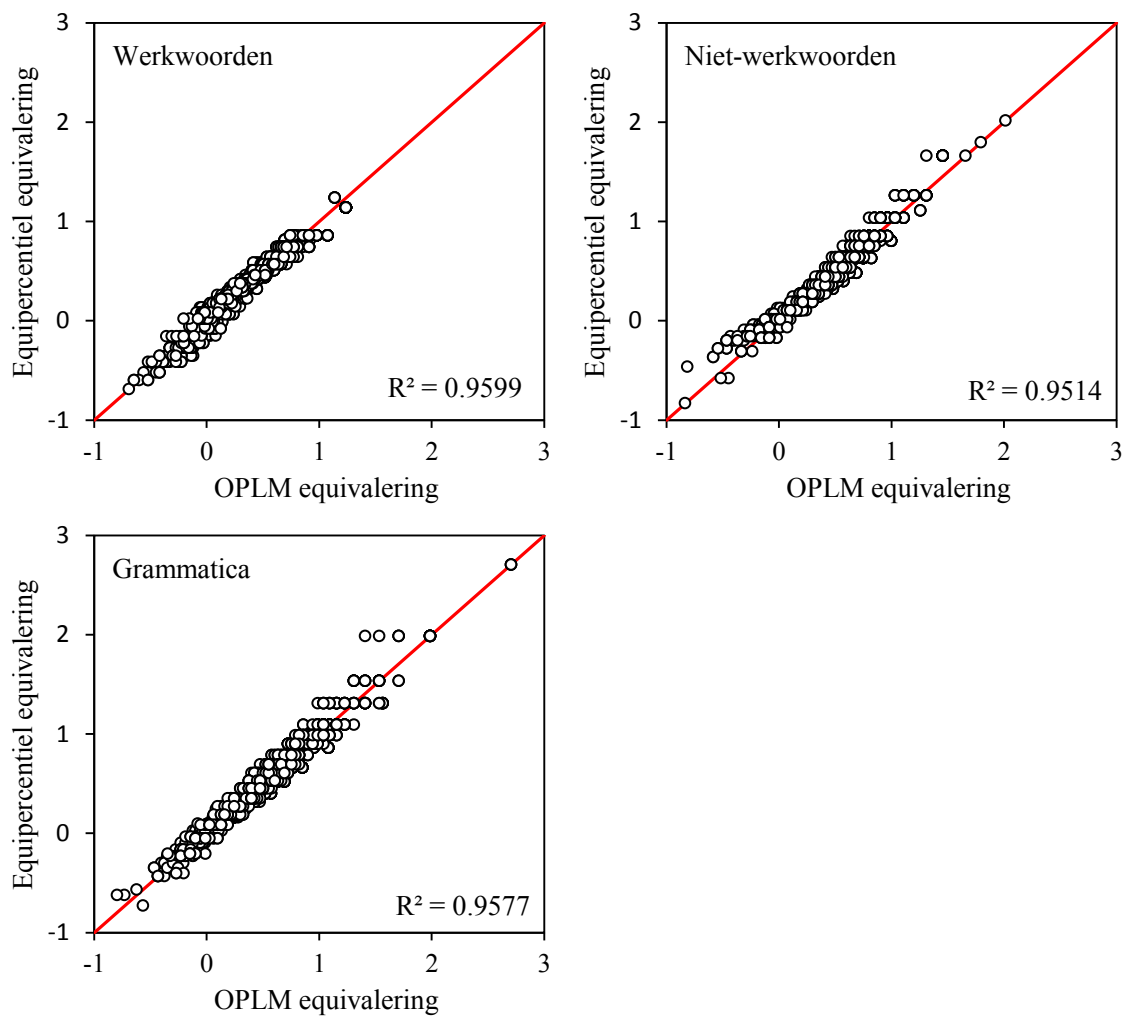
Het Cito Volgsysteem VO bevat naast een leestoets ook een toets taalverzorging. Er zijn in het derde leerjaar van het voortgezet onderwijs drie toetsversies beschikbaar. De eerste toets bevat 56 items en is bedoeld voor leerlingen in de basisberoepsgerichte leerweg van het VMBO. In het databestand is aan deze toets nummer 10 toegekend. De tweede toets (**BGL_VERS** = 11) bevat 60 items en is bedoeld voor leerlingen in de kaderberoepsgerichte, en gemengde en theoretische leerweg van het VMBO. De derde toets wordt afgenomen bij HAVO en VWO leerlingen (**BGL_VERS** = 12) en bevat net zoals de tweede toetsversie 60 items. De drie toetsversies liggen al op dezelfde meetschaal, maar wel op een andere dan die in COOL⁵⁻¹⁸ gebruikt wordt. Daarom is geprobeerd om met behulp van het OPLM een koppeling aan te brengen tussen de meetschalen. We zijn op exact dezelfde manier te werk gegaan als eerder bij begrijpend lezen. In paragraaf 4.8 is een uitgebreide toelichting hiervan te vinden.

Van de 1567 leerlingen van wie toetsgegevens waren ingestuurd, bleken er 17 de toets voor de basisberoepsgerichte leerweg van het VMBO gemaakt te hebben, 474 de toets voor de kaderberoepsgerichte, en gemengde en theoretische leerweg van het VMBO, en 910 de toets voor het

HAVO en het VWO. Dit betekent dat het ook nu niet mogelijk is om alle toetsversies van het Cito Volgsysteem VO te relateren aan COOL⁵⁻¹⁸. De eerste toets (**TVZ_VERS** = 10) is door te weinig leerlingen gemaakt om een zinvolle equivalering te kunnen uitvoeren. Bij de twee andere toetsen kon het OPLM wel zinvol geschat worden. Uit de statistische toetsen bleek dat het OPLM zowel bij werkwoorden als niet-werkwoorden en grammatica een adequate beschrijving geeft van de prestaties van de leerlingen: (grammatica) $R/c = 141$, $df = 152$, $p = .74$; (niet-werkwoorden) $R/c = 140$, $df = 152$, $p = .76$; (werkwoorden) $R/c = 226$, $df = 152$, $p < .01$. Het lijkt dan ook mogelijk om de scores die de leerlingen behaald hebben op de onderdelen van de toetsen taalverzorging te vertalen naar de meetschalen van COOL⁵⁻¹⁸.

Hoewel de equivalering op basis van het OPLM verdedigbaar lijkt, is een controle uitgevoerd via een (klassieke) equipercentielevivalering. Een equipercentielevivalering verschilt in opzet fundamenteel van een equivalering met behulp van het OPLM. Desalniettemin zou de equivalering vanwege de sterke inhoudelijke overeenkomsten tussen het Cito Volgsysteem VO en COOL⁵⁻¹⁸ wel een vergelijkbare uitkomst moeten opleveren. Bij de equipercentielevivalering hebben we het relatieve niveau van leerlingen bepaald in de totale groep met leerlingen die aan de afname van het Cito Volsysteem VO heeft deelgenomen ($N = 1384$). Er is dus geen onderscheid gemaakt in onderwijstypen. Figuur 5.5 laat per categorie zien hoe de equipercentielevivalering zich verhoudt tot de equivalering met het OPLM. We zien dat er zowel bij werkwoorden als niet-werkwoorden en grammatica sprake is van een sterke lineaire samenhang. De correlatie is afgerond steeds gelijk aan .98. Voor het bepalen van de geëquivalenteerde scores van de leerlingen maakt het dus weinig uit welke aanpak we kiezen. Wij hebben ervoor gekozen om gebruik te maken van de scores die voortvloeien uit de OPLM equivalering.

Net als eerder bij toetsversies 7, 8 en 9 bevat het databestand voor toetsversies 11 en 12 de ruwe totaalscore (d.w.z., het totaal op alle categorieën samen) en een composite score (of bankscore). Bij het berekenen van de composite scores voor de toetsen van het Cito Volgsysteem VO is te werk gegaan volgens de aanpak die is beschreven in paragraaf 5.6. De ruwe scores (**TVZ_RSCO**) hebben betrekking op toetsversies 11 en 12 en kunnen niet vergeleken worden met de ruwe scores van de andere toetsversies. Bij zowel toetsversie 11 als 12 is het maximum gelijk aan 60. Met behulp van de composite scores (**TVZ_BANK3**, max. = 100) kunnen wel alle toetsversies met elkaar vergeleken worden. Het heeft de voorkeur om bij analyses *binnen* een toetsversie gebruik te maken van de ruwe scores. Analyses over toetsversies of onderwijstypen heen vinden bij voorkeur plaats op basis van de composite scores die in het databestand zijn opgenomen.



Figuur 5.5. Equipercntielequivalering vergeleken met een OPLM equivalering

6 WISKUNDE

6.1 Materiaal

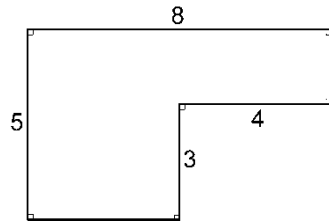
Bij de meting in 2014 zijn dezelfde wiskundetoetsen gebruikt als in 2011. In Zijsling et al. (2013) is een uitgebreide beschrijving van de inhoud van de toetsen te vinden. De toetsen zijn afgestemd op het *Referentiekader Doorlopende Leerlijnen Rekenen* (zie Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008). De vragen zijn onder te brengen in vier categorieën, namelijk *getallen*, *meten en meetkunde*, *verbanden* en *verhoudingen*. Hieronder staat voor elke categorie een voorbeeldvraag:

Getallen	Hoe schrijf je $\frac{31}{1000}$ als kommagetal?	
	A	0,31
	B	0,0031
	C	30,001
	D	0,031

Verhoudingen Een groepje van 5 mannen in een bedrijf koopt per maand samen één lot in de staatsloterij. Datzelfde doet een groepje van 8 vrouwen.
Er is elke maand één trekking. Dan ziet iedereen op welke lotnummers de prijzen zijn gevallen. Als er een prijs is gevallen op het lot van zo'n groepje, wordt de prijs onder de leden van de groep verdeeld, bij de mannen onder de 5 deelnemers, bij de vrouwen onder de 8 deelnemers.
In april 2009 viel een prijs van 100 000 euro op het lot dat de mannen gekocht hadden. Tegelijk viel een prijs van 200 000 euro op het lot dat de 8 vrouwen gekocht hadden. Bij deze loterij-uitslag kreeg elke man een bedrag, en elke vrouw een ander bedrag.
Het bedrag dat elke man kreeg, was

- A $\frac{4}{5}$ keer het bedrag dat elke vrouw kreeg
- B $\frac{5}{8}$ keer het bedrag dat elke vrouw kreeg
- C $\frac{5}{4}$ keer het bedrag dat elke vrouw kreeg
- D $\frac{8}{5}$ keer het bedrag dat elke vrouw kreeg

Meten en meetkunde



De hoeken in de figuur zijn alle recht.

De omtrek van de getekende figuur is

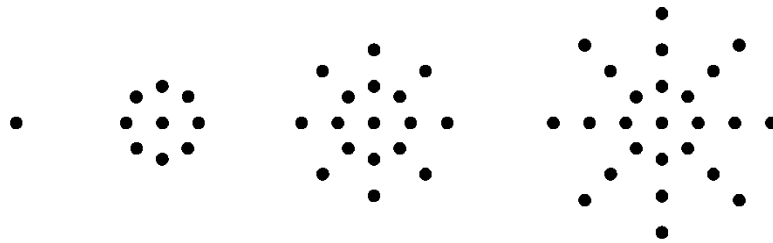
- A 20
- B 26
- C 28
- D 40

Verbanden

Deze opgave gaat over een rij figuren die bestaan uit stippen.

De figuren zijn volgens een vast patroon opgebouwd.

Hieronder staan de eerste vier figuren van de rij.



Figuur 1

Figuur 2

Figuur 3

Figuur 4

Het aantal stippen in de 10^e figuur is

- A 72
- B 73
- C 80
- D 81

6.2 Onderzoeksdesign

Tijdens de meting in 2014 zijn in totaal 60 verschillende items afgenomen. De items zijn afgenomen volgens een structureel onvolledig design. Figuur 6.1 geeft het design zoals dat gebruikt is tijdens de afname. Eerst zijn de items verdeeld in vier modules (A tot en met D, zie ook Tabel 6.1) met 15 items. Vervolgens zijn per onderwijstype twee modules geselecteerd die qua inhoud en moeilijkheidsgraad geschikt leken voor afname. VMBO leerlingen kregen afhankelijk van de leerweg modules A, B, of C voorgelegd en HAVO en VWO leerlingen kregen modules C en D voorgelegd. In het design is er voor gezorgd dat er zogenaamde ankermodules waren. Dat betekent dat een bepaalde groep leerlingen steeds een aantal items gemeenschappelijk had met een andere groep leerlingen. In het databestand geeft de variabele **WIS_VERS** aan welke combinatie van modules aan de leerling is voorgelegd.

Versie	Onderwijstype	Module			
		A	B	C	D
9	BBL				
10	KBL/GL/TL				
11	HAVO/VWO				

Figuur 6.1. *Afnamedesign wiskunde meting 3*

Een deel van de items die bij de meting in 2011 en 2014 is gebruikt, is ook al eerder gebruikt, namelijk bij de eerste meting van COOL⁵⁻¹⁸ in 2008 (zie Zijssling et al., 2009) en bij de laatste meting van VOCL'99 in 2002 (zie Kuyper & Van der Werf, 2005; Zijssling, Kuyper, Lubbers, & van der Werf, 2005). Door die overlap ontstaan er mogelijkheden om de verschillende metingen in de tijd op het gebied van wiskunde aan elkaar te relateren. Deze koppeling wordt echter in een aanvullend rapport beschreven (Keuning, Keizer-Mittelhaeuser, & Timmermans, in voorbereiding). Dit betekent dat de scores die in dit rapport beschreven worden alleen gebruikt kunnen worden om analyses te verrichten *binnen* de derde meting van COOL⁵⁻¹⁸ in VO-3. Vergelijkingen met de eerdere metingen van COOL⁵⁻¹⁸ of de laatste meting van VOCL'99 kunnen *niet* gemaakt worden op basis van dit rapport en het bijbehorende databestand.

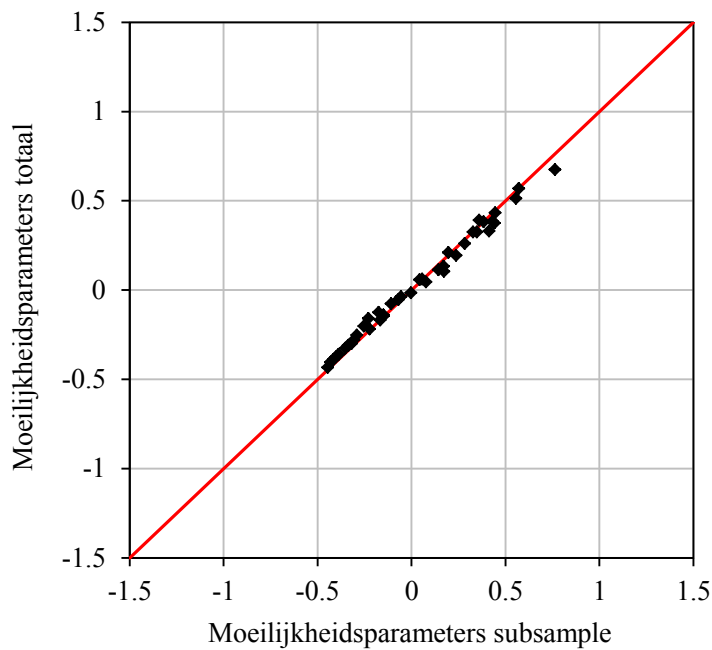
6.3 Meetmodel

De antwoorden van de leerlingen op de items zijn geanalyseerd met het OPLM voor dichtome items. Paragraaf 4.3 gaat in op de belangrijkste kenmerken en eigenschappen van het OPLM.

6.4 Kalibratie

Net zoals bij begrijpend lezen en taalverzorging varieerde het aantal waarnemingen bij wiskunde als gevolg van het afnamedesign sterk. De items in module A zijn bijvoorbeeld door 527 leerlingen gemaakt, terwijl de items in module C door 10.111 leerlingen zijn gemaakt. Hoewel het OPLM zonder problemen geschat kan worden met een dergelijke verdeling van waarnemingen, zal het beoordelen van de fit van het OPLM problematisch worden (zie ook paragraaf 4.4). Om problemen bij het beoordelen van de passing van het model te vermijden, is de analyse in twee stappen uitgevoerd. In de eerste stap is het OPLM geschat op basis van alle beschikbare waarnemingen. In een vervolgstap is het OPLM geschat op basis van een random *sample* uit de totale dataset waarin het aantal waarnemingen per item varieerde van minimaal 400 tot maximaal 800. Bij de beoordeling van de passing van het OPLM is in eerste instantie uitgegaan van de resultaten zoals die verkregen werden op basis van de *subsample*. Daarna is de samenhang tussen de schattingen op basis van de totale dataset en de schattingen op basis van de *subsample* in kaart gebracht. De uiteindelijke kalibratie is gebaseerd op de dataset met alle waarnemingen.

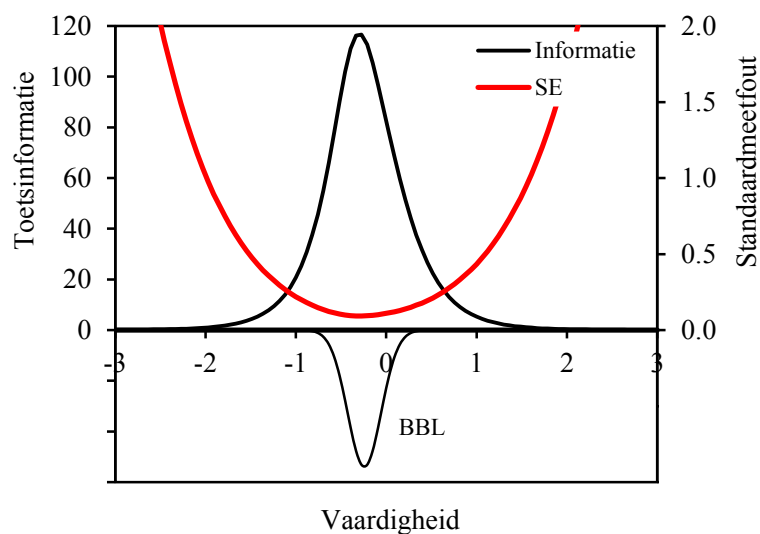
De kalibratie op basis van de *subsample* liet zien dat de prestaties van de leerlingen op de wiskunde items adequaat beschreven kunnen worden door het OPLM. Ten eerste bleek de verdeling van *p*-waarden (overschrijdingskansen) voor de *S*-toetsen bij elke categorie redelijk uniform verdeeld te zijn. Ten tweede was de verhouding tussen de *Rlc*-bijdrage en het aantal vrijheidsgraden acceptabel, $Rlc = 391$, $df = 286$, $p < .01$. De kalibraties op basis van de *subsample*, ten slotte, vertoonde grote gelijkenis met de kalibratie op basis van de dataset met alle waarnemingen. Figuur 6.2 laat duidelijk zien dat de relatie tussen de twee sets met moeilijkheidsparameters lineair is. Bij één item bleken de prestaties van leerlingen echter niet goed beschreven te kunnen worden door het OPLM (nummer 22 in versie 11). Dit item is daarom uit de schaal gelaten.



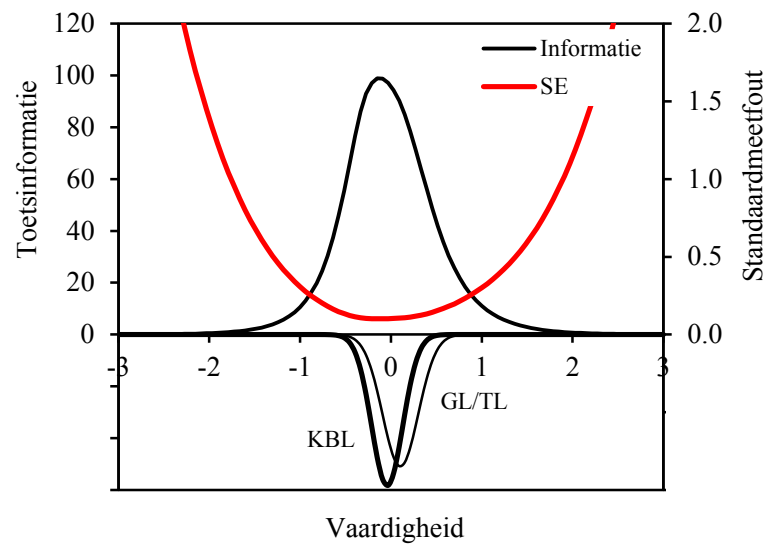
Figuur 6.2. *Relatie tussen moeilijkheidsparameters in twee verschillende kalibraties*

6.5 Meetnauwkeurigheid

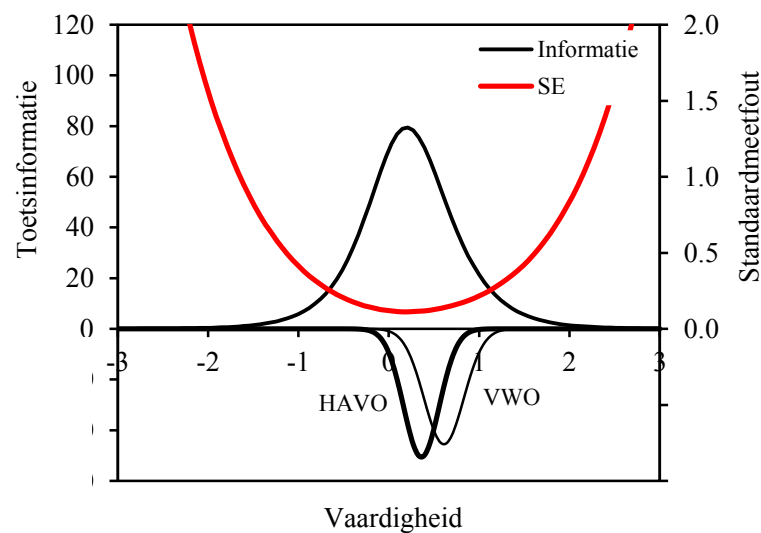
Figuren 6.3a, 6.3b en 6.3c geven voor elke toetsversie (zie Figuur 6.1 en **WIS_VERS** in het databestand) de toetsinformatiefunctie en de standaardmeetfout. Onder de x-as staan de kansdichtheidsfuncties (de ongewogen vaardigheidsverdelingen zoals geschat via *marginal maximum likelihood* met gefixeerde itemparameters) van de groepen leerlingen die de betreffende toets gemaakt hebben. De figuren laten zien dat toetsversies 9 en 10 optimaal aansloten bij het vaardigheidsniveau van de leerlingen die de toetsen maakten. De informatiefunctie vindt zijn piek namelijk steeds rond het gemiddelde. Toetsversie 11 paste goed bij het gemiddelde vaardigheidsniveau van HAVO-leerlingen, maar was wat aan de gemakkelijke kant voor de VWO-leerlingen.



Figuur 6.3a. *Toetsinformatiefunctie wiskundetoets versie 9*



Figuur 6.3b. Toetsinformatiefunctie wiskundetoets versie 10



Figuur 6.3c. Toetsinformatiefunctie wiskundetoets versie 11

Omdat er geen criteria beschikbaar zijn om toetsinformatiefuncties te beoordelen, is ook de klassieke betrouwbaarheid van de afgenomen wiskundetoetsen bepaald. Eerst is de *MAcc* coëfficiënt in de totale populatie bepaald onder de veronderstelling dat alle 59 geschaalde wiskunde items zijn afgenomen bij alle leerlingen. De *MAcc* coëfficiënt in de totale populatie voor alle 59 geschaalde items bleek gelijk aan .92. Vervolgens is met behulp van de Spearman-Brown-formule een schatting gemaakt van de betrouwbaarheid in geval k_t items zijn afgenomen. De betrouwbaarheid voor $k_t = 30$ is gelijk aan .85 (versie 9 en 10) en voor $k_t = 29$ is gelijk aan .85 (versie 11). De hiervoor genoemde toetslengtes wijken af van de toetslengtes die volgen uit Figuur 4.1, omdat één item niet in de meetschaal bleken te passen. De betrouwbaarheids-schattingen die verkregen worden na toepassing van de Spearman-Brown-formule zijn bovendien een onderschatting van de werkelijke betrouwbaarheid. Een toetsversie bestaat immers niet uit een *randomselectie* van items (de aanname waaronder de betrouwbaarheid is berekend), maar uit een selectie die *a priori* op basis van informatie over de moeilijkheidsgraad en het discriminerend vermogen van de items is gemaakt.

6.6 Toetsresultaten

In het databestand staan ruwe scores, gewogen scores, vaardigheidsscores en bankscores (zie paragraaf 4.6). De betreffende variabelen heten respectievelijk **WIS_RSCO**, **WIS_WSCO**, **WIS_THET** en **WIS_BANK3**. De ruwe en gewogen scores zijn niet vergelijkbaar over toetsversies. De bank- en vaardigheidsscores zijn dat wel. Bij het bepalen van de ruwe en gewogen scores zijn alleen de geschaalde items meegenomen. Dit betekent dat de maximale ruwe score voor versie 9 en 10 gelijk is aan 30 en voor versie 11 gelijk aan 29. De maximale gewogen scores zijn respectievelijk gelijk aan 117, 114 en 97. Het heeft de voorkeur om bij analyses *binnen* een onderwijstype (of toetsversie) gebruik te maken van de ruwe scores. Analyses over onderwijstypen heen vinden bij voorkeur plaats op basis van de bankscores die in het databestand zijn opgenomen.

6.7 Aanvullend onderzoek in Limburg

Net zoals bij begrijpend lezen is ook bij wiskunde samengewerkt met de onderzoekers van Inventaar 3VO. Dit betekent dat de wiskundevaardigheid van leerlingen gemeten is op basis van een korte toets die materiaal bevatte uit zowel COOL⁵⁻¹⁸ als Inventaar 3VO. Door items van COOL⁵⁻¹⁸ en Inventaar 3VO gezamenlijk in één toets af te nemen (de zogeheten “Limburg-versie”), wordt het mogelijk om de toetsscores die de leerlingen behalen te vertalen naar de meetschaal die in COOL⁵⁻¹⁸ gebruikt wordt. Andersom is ook mogelijk; de toetsscores die behaald zijn op COOL⁵⁻¹⁸, zijn ook te vertalen naar de meetschalen van Inventaar 3VO.

De “Limburg-versie” bevatte in totaal 19 items, waarvan er 9 uit COOL⁵⁻¹⁸ kwamen. De betreffende items zijn op papier afgenomen. De overige 10 items kwamen uit Inventaar 3VO. Omdat Inventaar 3VO gebruikmaakt van digitale toetsen zijn deze items via de computer afgenomen. De volgorde van afname is gehusseld. Dit betekent dat de ene groep leerlingen tijdens de afname eerst het papieren gedeelte maakte en daarna doorging met het digitale gedeelte (versie 12 in het databestand), terwijl de andere groep leerlingen in omgekeerde volgorde te werk ging (versie 13 in het databestand). De “Limburg-versie” is gebruikt in alle onderwijstypen. De leerlingen in de basisberoepsgerichte leerweg van het VMBO kregen dus dezelfde items voorgelegd als leerlingen in het HAVO en het VWO. Omdat vooraf niet geheel duidelijk was hoe de items zouden functioneren in de verschillende onderwijstypen, en of de afnameconditie (papier of digitaal) effect zou hebben op de prestaties van leerlingen, zijn voorafgaand aan de equivalering enkele verkennende analyses verricht.

Eerst is per onderwijstype een toets- en itemanalyse uitgevoerd. Tabel 6.1 bevat de belangrijkste resultaten. De leerlingen die niet alle items gemaakt hadden, zijn buiten de analyse gelaten. Binnen elk van de onderwijstypen staat in de eerste kolom het aantal leerlingen. In de tweede en derde kolom staat respectievelijk de p -waarde en de r_{it} -waarde. In Tabel 6.1 is te zien dat de p -waarde bij de meeste items acceptabel is, zij het dat de items voor de leerlingen in de basis- en kaderberoepsgerichte leerweg van het VMBO soms aan de moeilijke kant waren (p -waarde $< .40$) en voor de leerlingen in het VWO soms aan de gemakkelijke kant (p -waarde $> .90$). De verschillen tussen de VMBO gemengde en theoretische leerweg en het HAVO zijn gering. De r_{it} -waarden liggen veelal boven de .20 en zijn daarmee acceptabel te noemen. Er zijn echter uitzonderingen, zoals de r_{it} -waarde van item 4 in de twee laagste leerwegen van het VMBO. Over het algemeen functioneren de items met een r_{it} -waarde $< .20$ verschillend over onderwijstypen. Veldhuijzen, Goldebelt en Sanders (1993) geven 95% betrouwbaarheidsintervallen voor r_{it} -waarden. Bij een gering aantal leerlingen (i.e., 100 of 200) en een geobserveerde r_{it} -waarde van .20 ligt de r_{it} -waarde bij herhaling in 95 procent van de gevallen in het interval [.00; .40]. Hoewel een dergelijk item mogelijk niet

in grote mate bijdraagt aan het meten van de wiskundevaardigheid binnen de desbetreffende populatie, is er in het licht van dit grote interval ook niet direct reden om het item uit de vervolganalyse te laten. Dit is dan ook niet gedaan.

Tabel 6.1

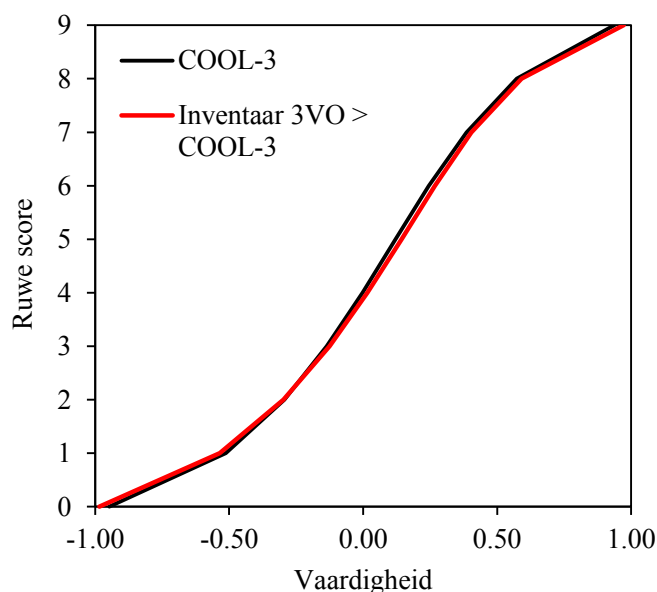
Toets- en itemanalyse wiskundetoets versies 12 en 13

Item	BBL + KBL			GL + TL			HAVO			VWO		
	<i>n</i>	<i>p</i>	<i>r_{it}</i>	<i>n</i>	<i>p</i>	<i>r_{it}</i>	<i>n</i>	<i>p</i>	<i>r_{it}</i>	<i>n</i>	<i>p</i>	<i>r_{it}</i>
1	113	.60	.30	203	.71	.34	546	.72	.25	374	.85	.33
2	113	.64	.44	203	.83	.35	546	.84	.29	374	.94	.36
3	113	.51	.39	203	.82	.37	546	.85	.29	374	.94	.17
4	113	.52	.17	203	.76	.42	546	.81	.28	374	.90	.33
5	113	.53	.31	203	.79	.36	546	.82	.35	374	.90	.26
6	113	.45	.26	203	.49	.21	546	.71	.29	374	.82	.43
7	113	.25	.26	203	.49	.39	546	.68	.31	374	.91	.20
8	113	.27	.36	203	.65	.37	546	.62	.33	374	.76	.31
9	113	.47	.14	203	.75	.35	546	.75	.36	374	.90	.23
10	113	.67	.33	203	.76	.23	546	.69	.36	374	.88	.36
11	113	.50	.48	203	.78	.33	546	.73	.49	374	.88	.27
12	113	.54	.10	203	.61	.17	546	.55	.22	374	.63	.36
13	113	.69	.23	203	.93	.27	546	.83	.32	374	.96	.11
14	113	.27	.34	203	.61	.27	546	.61	.42	374	.82	.38
15	113	.41	.28	203	.80	.45	546	.79	.46	374	.97	.30
16	113	.50	.35	203	.80	.50	546	.80	.41	374	.96	.22
17	113	.31	.16	203	.40	.32	546	.50	.30	374	.70	.40
18	113	.42	.50	203	.66	.48	546	.65	.41	374	.83	.33
19	113	.25	.40	203	.56	.33	546	.55	.49	374	.83	.42

In een vervolgstap zijn toetsversies 12 en 13 geanalyseerd met behulp van het OPLM. Deze analyse had als doel om vast te stellen of (a) de papieren en digitale items ondergebracht kunnen worden in één meetschaal en (b) de afnameconditie effect heeft op de prestaties van leerlingen. De prestaties van de leerlingen bleken adequaat beschreven te kunnen worden door het OPLM. Ten eerste bleek de verdeling van *p*-waarden (overschrijdingskansen) voor de *S*-toetsen bij elke categorie redelijk uniform verdeeld te zijn. Ten tweede was de verhouding tussen de *R*/*I*-bijdrage en het aantal vrijheidsgraden acceptabel, *R*/*I* = 152, *df* = 126, *p* = .05. Dit betekent dat de papieren en digitale items hoogstwaarschijnlijk een beroep doen op dezelfde onderliggende vaardigheid en daarmee ondergebracht kunnen worden in één meetschaal. Zodra de papieren en digitale items op één meetschaal zijn geplaatst, wordt het ook mogelijk om de twee afnamecondities met elkaar te vergelijken. Door middel van een gepaarde *t*-toets is nagegaan in hoeverre leerlingen in de ene conditie een andere vaardigheid laten zien dan in de andere conditie. Als we de prestaties van de leerlingen met behulp van het OPLM corrigeren voor de moeilijkheidsgraad van de items, blijken leerlingen in de papieren (*M* = 0.339, *SD* = 0.366) afnameconditie niet significant hoger of lager te scoren dan in de digitale (*M* = 0.328, *SD* = 0.293) afnameconditie; *t*(1342) = 1.142, *p* = 0.254. Dit resultaat bevestigt nog eens dat het digitale gedeelte equivalent is aan het papieren gedeelte en er daarom met één meetschaal gewerkt kan worden.

Na deze verkennende analyses is de meetschaal voor Limburg gerelateerd aan de meetschaal voor COOL-3. Er zijn verschillende manieren om dat te doen. Een eerste mogelijkheid is om de parameters die eerder geschat zijn voor de items van toetsversies 8, 9 en 10 (zie paragraaf 4.4) te fixeren en de items van toetsversies 11 en 12 er als het ware “bij te schalen”. Deze aanpak bleek, net als bij begrijpend lezen, niet optimaal te werken. Sommige item karakteristieke curven bleken namelijk niet te passen bij het antwoordgedrag van de Inventaar 3VO leerlingen. Als we dit zouden negeren bij het schatten van de vaardigheid van leerlingen bestaat het risico op onder- of overschatting. Daarom is gekozen voor een alternatief.

Om in de equivalering enigszins rekening te kunnen houden met het “afwijkende” antwoordgedrag van de Inventaar 3VO leerlingen is wederom gebruikgemaakt van de zogeheten *proficiencytransformation* procedure (zie paragraaf 4.7 voor een toelichting). Figuur 6.4 laat zien hoe de procedure uitpakt bij de koppeling van Inventaar 3VO aan COOL-3. De zwarte lijn representeert de relatie tussen de vaardigheidsscores en de (ongewogen) ruwe scores zoals die geldt in COOL-3. De rode lijn representeert dezelfde relatie, maar dan voor Inventaar 3VO ná toepassing van de hiervoor beschreven transformatie. Bij het construeren van de lijnen zijn vanzelfsprekend alleen de overlappende items meegenomen. De rode lijn zou exact over de zwarte lijn heen liggen als de Inventaar 3VO leerlingen hetzelfde antwoordgedrag vertonen als de COOL-3 leerlingen. De rode lijn kan verder van de zwarte lijn af komen te liggen naarmate de verschillen in antwoordgedrag groter zijn. De rode lijn vertoont grote gelijkheid met de zwarte lijn en hiermee lijkt de *proficiencytransformation* procedure een verdedigbare koppeling tussen COOL-3 en Inventaar 3VO op te leveren. Op basis van deze transformatie zijn dan ook de (geëquivalenteerde) scores van de leerlingen bepaald.



Figuur 6.4. *Vergelijking van Inventaar 3VO en COOL-3 op basis van de toets karakteristieke curve*

Net als eerder bij toetsversies 9, 10 en 11 bevat het databestand voor toetsversies 12 en 13 de ruwe scores (**WIS_RSCO**, max =19), de gewogen scores (**WIS_WSCO**, max. = 73), de vaardigheidsscores (**WIS_THET**) en de bankscores (**WIS_BANK3**, max. = 100). De ruwe en gewogen scores hebben betrekking op toetsversies 12 en 13 en kunnen niet vergeleken worden met de ruwe en gewogen scores van de andere toetsversies. De vaardigheidsscores zijn in eerste instantie geschat op de Inventaar 3VO meetschaal. Vervolgens zijn de vaardigheidsscores via de eerder beschreven lineaire transformatie

gerelateerd aan de COOL-3 meetschaal, zodat het mogelijk wordt om vergelijkingen te maken tussen Inventaar 3VO en COOL-3. De bankscores zijn berekend op basis van de lineair getransformeerde vaardigheidsscores en de itemparameters zoals die zijn geschat op basis van de COOL-3 data. De 10 items van Inventaar 3VO zijn niet meegenomen bij het berekenen van de bankscores.

Hoewel Figuur 6.4 erop wijst dat de koppeling die gemaakt is tussen Inventaar 3VO en COOL-3 adequaat is, moeten de bank- en vaardigheidsscores zoals die berekend zijn voor toetsversies 12 en 13 wel met de nodige voorzichtigheid gebruikt worden. De aandachtspunten die eerder van toepassing waren bij begrijpend lezen (zie paragraaf 4.7) gelden ook bij wiskunde. Dit betekent het volgende:

1. De toets die in Limburg gebruikt is, was tamelijk kort en de overlap met COOL-3 was beperkt. Uiteindelijk is de koppeling tussen Inventaar 3VO en COOL-3 gebaseerd op 9 items. Een groter anker leidt in de regel tot een stabielere (of nauwkeurigere) equivalering, maar dat was in het onderhavige onderzoek om praktische redenen niet te realiseren.
2. Het antwoordgedrag van de leerlingen die deelnamen aan Inventaar 3VO week soms af van het antwoordgedrag dat werd waargenomen in COOL-3. Door voor Inventaar 3VO een aparte meetschaal te construeren en deze pas in een vervolgstap via de zogeheten *proficiencytransformation* procedure te relateren aan de meetschaal van COOL-3 konden we het risico op bias in de vaardigheidsscores beperken. De lineaire transformatie doet immers geen afbreuk aan de kwaliteit van de vaardigheidsscores. Dit geldt echter alleen voor analyses binnen Inventaar 3VO. Als een vergelijking gemaakt wordt tussen Inventaar 3VO en COOL-3 speelt de lineaire transformatie wel een rol. De transformatie lijkt verdedigbaar, maar uiteindelijk is het vanwege de waargenomen verschillen in antwoordgedrag niet geheel zeker of de vergelijking tussen Inventaar 3VO en COOL-3 ook daadwerkelijk adequaat te maken is.
3. Bij het berekenen van de bankscores is gebruikgemaakt van de lineair getransformeerde vaardigheidsscores en de itemparameters van COOL-3. De vaardigheidsscores zijn gebaseerd op 19 items. De betrouwbaarheid is in het licht van de toetslengte goed te noemen (.77) en ook uit de toetsinformatiefunctie blijkt dat toetsversies 12 en 13 goed aansloten bij het vaardigheidsniveau van de leerlingen die de toetsen maakten. Alleen voor de (betere) leerlingen in het HAVO en VWO was de toets aan de gemakkelijke kant en zijn de vaardigheidsscores behept met een grotere meetfout. Dit heeft uiteraard ook zijn weerslag op de bankscores, maar omdat analyses doorgaans op groepsniveau plaatsvinden is de iets grotere meetfout aan de bovenkant van de meetschaal geen groot probleem. Wel blijft staan dat de bankscores gebaseerd zijn op de itemparameters van COOL-3 en dat daarmee verondersteld is dat deze geldig zijn voor Inventaar 3VO. Zoals al eerder is aangegeven, is deze veronderstelling alleen te controleren voor de 9 ankeritems. Of deze veronderstelling ook te verdedigen is voor de overige items is niet te onderzoeken.

6.8 Koppeling met het Cito volgsysteem VO

Het Cito Volgsysteem VO bevat voor het derde leerjaar drie verschillende wiskundetoetsen. De toetsen zijn bedoeld voor verschillende onderwijstypen. De eerste toets bevat 93 items en is bedoeld voor leerlingen in de basisberoepsgerichte leerweg van het VMBO. In het databestand is aan deze toets nummer 14 toegekend. De tweede toets (**WIS_VERS** = 15) bevat 72 items en is bedoeld voor leerlingen in de kaderberoepsgerichte, en gemengde en theoretische leerweg van het VMBO. De derde toets wordt afgenomen bij HAVO en VWO leerlingen (**WIS_VERS** = 16) en bevat 75 items. De items in de wiskundetoetsen liggen niet op dezelfde meetschaal als de items van COOL⁵⁻¹⁸. We passen daarom een equivaleringsprocedure toe om het Cito Volgsysteem VO te relateren aan COOL⁵⁻¹⁸. Vaak vindt een equivalering plaats op basis van een set gemeenschappelijke items (zie ook paragraaf 4.2), maar deze

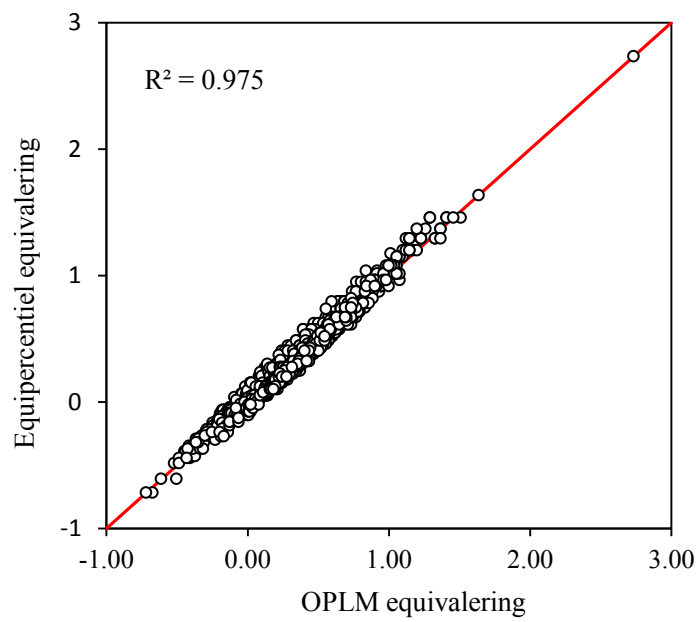
aanpak was niet toe te passen, omdat het Cito Volgsysteem VO geen (of hooguit enkele) items bevat die ook in COOL⁵⁻¹⁸ gebruikt worden.

Indien items (of toetsen) niet op dezelfde meetschaal geplaatst kunnen worden via gemeenschappelijke items of leerlingen, is een equivalering alleen mogelijk onder een aantal aannames. In het onderhavige geval is verondersteld dat de wiskundetoetsen van COOL⁵⁻¹⁸ en het Cito Volgsysteem VO gemaakt zijn door statistisch equivalente groepen. Er is, met andere woorden, aangenomen dat de leerlingen die de volgsysteemtoetsen maakten uit dezelfde onderliggende vaardigheidsverdeling afkomstig zijn als de leerlingen die de COOL⁵⁻¹⁸-toetsen maakten. Deze aanname is eerder ook bij begrijpend lezen gemaakt (zie paragraaf 4.8). Onder aanname van statistisch equivalente groepen is het mogelijk om het OPLM te schatten met behulp van *marginal maximum likelihood*(MML). We houden bij het schatten van de itemparameters dan rekening met de groep leerlingen die een toetsboekje maakt en doen daarnaast een veronderstelling over de verdeling van de vaardigheid in de populatie. Wij hebben aangenomen dat de vaardigheid in de populatie normaal verdeeld is.

In totaal hebben 7 scholen van 1567 leerlingen toetsgegevens ingestuurd. Sommige leerlingen bleken geen wiskundetoets te hebben gemaakt. Uiteindelijk konden we per toetsversie beschikken over de volgende hoeveelheid gegevens: (WIS_VERS 14) 17 leerlingen, (WIS_VERS 15) 541 leerlingen, en (WIS_VERS 16) 910 leerlingen. Het is opmerkelijk dat de toetsversie die bedoeld is voor de basisberoepsgerichte leerweg van het VMBO slechts door 17 leerlingen is gemaakt. De meeste leerlingen in de basisberoepsgerichte leerweg bleken een versie van een hoger niveau gemaakt te hebben. Met 17 waarnemingen is er geen stabiele basis om de toets te relateren aan COOL⁵⁻¹⁸. Er is dan ook besloten om daarvan af te zien. Bij de twee andere toetsversies hadden we wel voldoende gegevens en is het OPLM geschat met behulp van MML. Uit de statistische toetsen bleek dat het OPLM een adequate beschrijving geeft van de prestaties van de leerlingen: $R^2 = 465.59$, $df = 580$, $p = .99$. Het lijkt dan ook zonder meer mogelijk om de scores die de leerlingen behaald hebben op de wiskundetoetsen van het Cito Volgsysteem VO met behulp van dit model te vertalen naar de meetschaal van COOL⁵⁻¹⁸.

Hoewel de equivalering op basis van het OPLM net als eerder bij begrijpend lezen verdedigbaar lijkt, is de equivalering gecontroleerd op basis van een (klassieke) equipercentiequivalering zonder smoothing (zie paragraaf 4.8 voor een toelichting). Bij de equipercentiequivalering is geen onderscheid gemaakt in onderwijstypen. Dit betekent dat het relatieve niveau van leerlingen is bepaald in de totale groep met leerlingen die toetsversie 15 of 16 hebben gemaakt. Figuur 6.5 laat zien hoe de equipercentiequivalering zich verhoudt tot de equivalering met het OPLM. We zien dat er sprake is van een sterke lineaire samenhang, $r = .98$. Voor het bepalen van de geëquivalerde scores van de leerlingen maakt het dus weinig uit welke aanpak we kiezen. Wij hebben ervoor gekozen om gebruik te maken van de scores die voortvloeien uit de OPLM equivalering.

Net als bij eerdere toetsversies bevat het databestand voor toetsversies 15 en 16 de ruwe scores (**WIS_RSCO**, max. versie 15 = 72 / max. versie 16 = 75), de gewogen scores (**WIS_WSCO**, max. versie 15 = 269 / max. versie 16 = 276), de vaardigheidsscores (**WIS_THET**) en de bankscores (**WIS_BANK3**, max. = 100). De ruwe en gewogen scores hebben betrekking op toetsversies 15 en 16 en kunnen niet vergeleken worden met de ruwe en gewogen scores van de andere toetsversies. Op basis van de bank- en vaardigheidsscores kan die vergelijking wel gemaakt worden. Het heeft de voorkeur om bij analyses *binnen* een toetsversie gebruik te maken van de ruwe scores. Analyses over toetsversies of onderwijstypen heen vinden bij voorkeur plaats op basis van de bankscores die in het databestand zijn opgenomen.



Figuur 6.5. *Equipercntielequivalering vergeleken met een OPLM equivalering*

7 ENGELS

7.1 Materiaal

Bij de meting in 2014 zijn voor Engels dezelfde toetsen afgenomen als in 2011 en 2008. In de toets werden vragen gesteld over een aantal korte teksten. Leerlingen konden bij de beantwoording van de vragen kiezen uit drie of vier alternatieven en het antwoord moest aangestreept worden op een antwoordblad. Hieronder staat een voorbeelditem:

Voorbeeld **Waarom is deze oproep geplaatst?**

- A Om aan originele ideeën voor films te komen door middel van een filmprijsvraag.
- B Om deelnemers te vinden voor een filmquiz, waarbij bekende sterren inde jury zitten.
- C Om jong talent de gelegenheid te geven de filmwereld te leren kennen.
- D Om kandidaten voor de functie van filmrecensent op te roepen hun vaardigheid te demonstreren.

**See movies for free, meet the stars
and inflict your opinion on thousands!**

The Observer has linked up with Radio One's Clingfilm programme, presented by Wendy Lloyd and Mark Kermode to find a talented new film critic. If you are aged between 18–26 (born on or after 30.6.68), you can win an all expenses paid long weekend at the London Film Festival and the chance to review for The Observer and Radio One.

The competition will be judged by actor Richard E Grant, London Film Festival director Sheila Whitaker, The Observer's film critic Philip French and arts editor Jane Ferguson, and Radio One film critic Mark Kermode.

To enter, send an 800 word review of a film or group of films you have seen recently, with a one-minute cassette recording of you reading part of the review to: The Observer, Young Film Critics Competition, 119 Farringdon Road, London EC1R 3ER. Closing date September 15. (Please remember to label your cassette and announce yourself on the recording.)

THE OBSERVER

7.2 Onderzoekdesign

Tijdens de meting in 2014 zijn net zoals bij de eerdere meting van COOL⁵⁻¹⁸ 44 items afgenomen. De items zijn afgenomen volgens een structureel onvolledig design. Figuur 7.1 geeft het design weer zoals dat gebruikt is tijdens de afname. Eerst zijn de items verdeeld in vier modules met een wisselend aantal items. In modules A tot en met C zaten 10 items en in module D zaten 14 items. Vervolgens zijn per

onderwijstype twee modules geselecteerd die qua inhoud en moeilijkheid geschikt leken voor afname. VMBO leerlingen kregen afhankelijk van de leerweg modules A, B, of C voorgelegd en HAVO en VWO leerlingen kregen modules C en D voorgelegd. In het design is er voor gezorgd dat er zogenaamde ankermodules waren. Dat betekent dat een bepaalde groep leerlingen steeds een aantal items gemeenschappelijk had met een andere groep leerlingen. In het databestand geeft de variabele **ENG_VERS** aan welke combinatie van modules aan de leerling is voorgelegd.

Versie	Onderwijstype	Module			
		A	B	C	D
7	BBL				
8	KBL/GL/TL				
9	HAVO/VWO				

Figuur 7.1. *Afnamedesign Engels meting 3*

Zoals in paragraaf 7.1 al is opgemerkt, zijn bij de tweede meting in 2014 exact dezelfde toetsen gebruikt als bij de metingen in 2008 en 2011 (zie Zijssling et al., 2009; 2013). De scores die door de leerlingen behaald zijn bij de verschillende metingen kunnen echter *niet* met elkaar vergeleken worden, omdat er in COOL-3 voor gekozen is om de itemantwoorden van de leerlingen opnieuw te analyseren. Het gevolg van deze keuze is dat de gewogen scores en de vaardigheidsscores voor COOL-3 voortvloeien uit een ander meetmodel dan in COOL-1 en COOL-2 en er geen basis is voor het maken van een vergelijking tussen de metingen. Ditzelfde geldt voor de ruwe scores, omdat sommige items in COOL-3 vanwege misfit niet zijn meegenomen in de berekening van de ruwe scores. Vanzelfsprekend zijn er wel mogelijkheden om de metingen aan elkaar te relateren. Deze koppeling wordt echter niet in dit rapport, maar in een aanvullend rapport beschreven (Keuning, Keizer-Mittelhaeuser, & Timmermans, in voorbereiding). De scores die in dit rapport beschreven worden, kunnen dus alleen gebruikt worden om analyses te verrichten *binnen* de derde meting van COOL⁵⁻¹⁸ in VO-3. Vergelijkingen met de eerdere metingen van COOL⁵⁻¹⁸ kunnen *niet* gemaakt worden op basis van dit rapport en het bijbehorende databestand. Om dit nog eens extra te benadrukken worden in het databestand de versienummers 7, 8 en 9 gebruikt in plaats van de eerdere versienummers 1, 2 en 3 (2008), en 4, 5 en 6 (2011).

7.3 Meetmodel

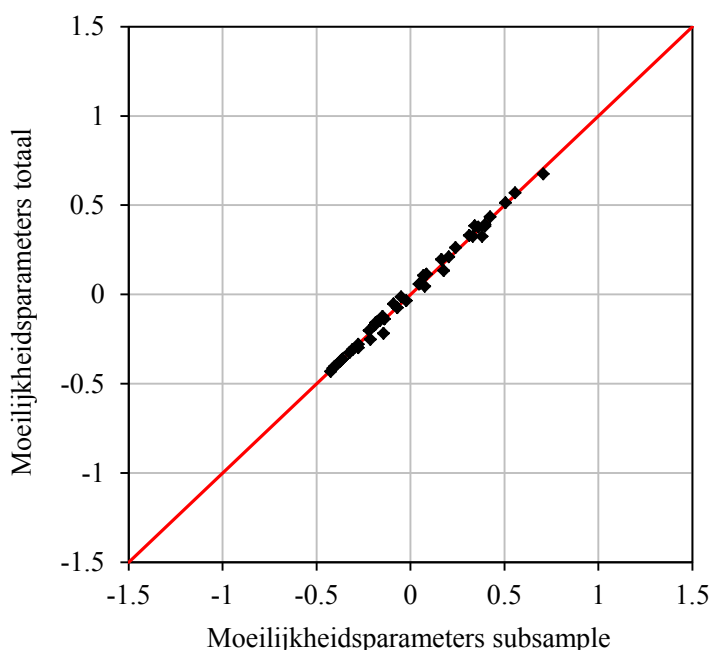
De antwoorden van de leerlingen op de items bij het onderdeel Engels zijn geanalyseerd met het OPLM voor dichtome items. Paragraaf 4.3 gaat in op de belangrijkste kenmerken en eigenschappen van het OPLM.

7.4 Kalibratie

Bij het onderdeel Engels varieerde het aantal waarnemingen als gevolg van het afnamedesign sterk. De items in module A zijn bijvoorbeeld door 274 leerlingen gemaakt, terwijl de items in module C door 7412 leerlingen zijn gemaakt. Met een aantal van 274 leerlingen is het onzeker of het OPLM adequaat te schatten is. Het is met name problematisch om de itemgewichten, a_j , te bepalen. In de regel wordt een leerlingaantal van 300 als ondergrens gehanteerd. Om problemen bij het schatten van het OPLM te vermijden, is er voor gekozen om de itemgewichten niet alleen te baseren op de data die verzameld zijn bij COOL-3, maar ook

op de data die bij eerdere metingen verzameld zijn. Bij het schatten van de moeilijkheidsparameters is wel alleen de data van de meest recente meting gebruikt. Verder is dezelfde werkwijze gevolgd als eerder bij begrijpend lezen, taalverzorging en wiskunde. Dit betekent dat de kalibratie in twee stappen is uitgevoerd. In de eerste stap is het OPLM geschat op basis van alle beschikbare waarnemingen. In een vervolgstap is het OPLM geschat op basis van een random *sample* uit de totale dataset waarin het aantal waarnemingen per item varieerde van minimaal 274 tot maximaal 800. Bij de beoordeling van de passing van het OPLM is in eerste instantie uitgegaan van de resultaten zoals die verkregen werden op basis van de *subsample*. Daarna is de samenhang tussen de schattingen op basis van de totale dataset en de schattingen op basis van de *subsample* in kaart gebracht. De uiteindelijke kalibratie is gebaseerd op de dataset met alle waarnemingen.

De kalibratie op basis van de *subsample* liet zien dat de prestaties van de leerlingen op de items adequaat beschreven kunnen worden door het OPLM. Ten eerste bleek de verdeling van *p*-waarden (overschrijdingskansen) voor de *S*-toetsen bij elke categorie redelijk uniform verdeeld te zijn. Ten tweede was de verhouding tussen de *Rlc*-bijdrage en het aantal vrijheidsgraden acceptabel, $Rlc = 352$, $df = 197$, $p < .01$. De kalibraties op basis van de *subsample*, ten slotte, vertoonde grote gelijkheid met de kalibratie op basis van de dataset met alle waarnemingen. Figuur 7.2 laat duidelijk zien dat de relatie tussen de twee sets met moeilijkheidsparameters lineair is. Bij één item bleek de prestatie van leerlingen niet goed beschreven te kunnen worden door het OPLM (nummer 17 in versie 8). Dit item is bij deze toetsversie daarom uit de schaal gelaten.

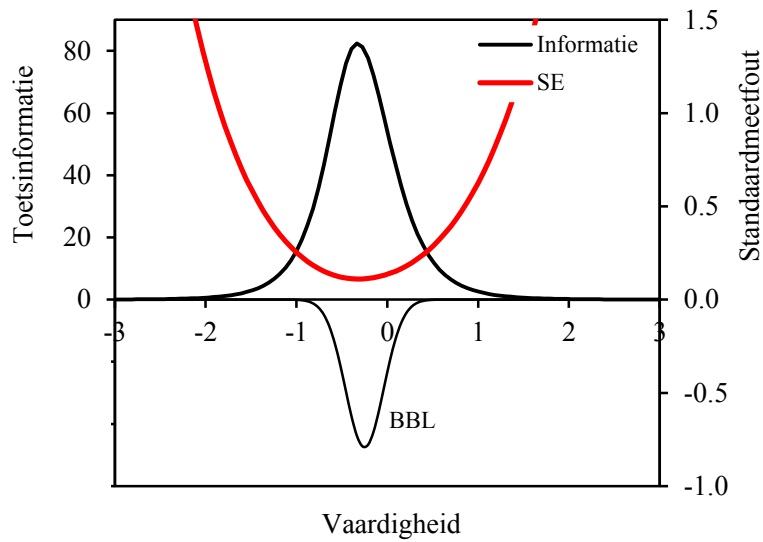


Figuur 7.2. Relatie tussen moeilijkheidsparameters in twee verschillende kalibraties

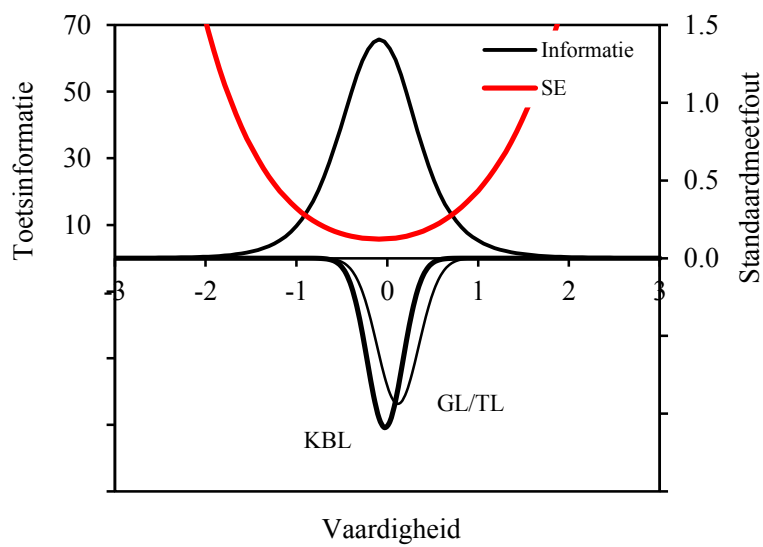
7.5 Meetnauwkeurigheid

Figuren 7.3a, 7.3b en 7.3c geven voor elke toetsversie (zie Figuur 7.1 en **ENG_VERS** in het databestand) de toetsinformatiefunctie en de standaardmeetfout. Onder de *x*-as staan de kansdichtheidsfuncties (de ongewogen vaardigheidsverdelingen zoals geschat via *marginal maximum likelihood* met gefixeerde itemparameters) van de groepen leerlingen die de betreffende toets gemaakt hebben. De figuren laten zien dat de items in de verschillende versies van de toets Engels over het algemeen goed pasten bij het vaardigheidsniveau in de doelgroep. Steeds bereikt de standaardmeetfout zijn minimum in het

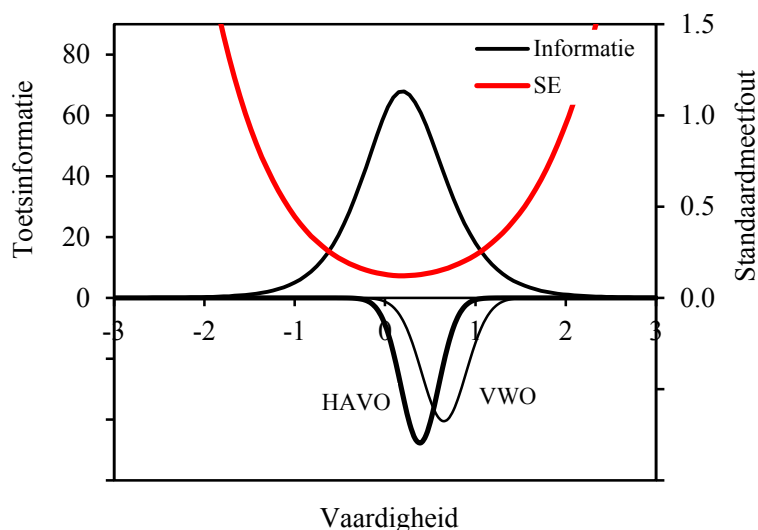
vaardigheidsgebied van de doelgroep. Voor VWO-leerlingen lijkt de toets wat aan de gemakkelijke kant te zijn geweest.



Figuur 7.3a. *Toetsinformatiefuncties toets Engels versie 7*



Figuur 7.3b. *Toetsinformatiefuncties toets Engels versie 8*



Figuur 7.3c. Toetsinformatiefuncties toets Engels versie 9

Zoals eerder aangegeven is het enigszins problematisch om een toetsinformatiefunctie te beoordelen. Daarom is ook de klassieke betrouwbaarheidscoëfficiënt ($MAcc$) bepaald. Eerst is de $MAcc$ coëfficiënt in de totale populatie bepaald onder de veronderstelling dat alle 44 geschaalde items zijn afgenomen bij alle leerlingen. De $MAcc$ coëfficiënt bleek gelijk aan .88. Vervolgens is met behulp van de Spearman-Brown-formule een schatting gemaakt van de betrouwbaarheid in geval k_t items zijn afgenomen. De betrouwbaarheid voor $k_t = 20$ is gelijk aan .82 (toetsversie 7), voor $k_t = 19$ is de betrouwbaarheid gelijk aan .82 (toetsversie 8), en voor $k_t = 24$ (toetsversie 9) is de betrouwbaarheid gelijk aan .85. De hiervoor genoemde toetslengtes wijken af van de toetslengtes die volgen uit Figuur 7.1, omdat één item niet in de schaal bleken te passen voor toetsversie 8. De betrouwbaarheidsschattingen die verkregen worden na toepassing van de Spearman-Brown-formule zijn bovendien een onderschatting van de werkelijke betrouwbaarheid. Een toetsversie bestaat immers niet uit een *random* selectie van items (de aanname waaronder de betrouwbaarheid is berekend), maar uit een selectie die *a priori* op basis van informatie over de moeilijkheidsgraad en het discriminerend vermogen van de items is gemaakt.

7.6 Toetsresultaten

In het databestand staan ruwe scores, gewogen scores, vaardigheidsscores en bankscores (zie paragraaf 4.6). De betreffende variabelen heten respectievelijk **ENG_RSCO**, **ENG_WSCO**, **ENG_THET** en **ENG_BANK3**. De ruwe en gewogen scores zijn niet vergelijkbaar over toetsversies. De bank- en vaardigheidsscores zijn dat wel. Bij het bepalen van de ruwe en gewogen scores zijn alleen de geschaalde items meegenomen. Dit betekent dat de maximale ruwe score voor versie 7 gelijk is aan 20, voor versie 8 gelijk aan 19 en voor versie 9 gelijk aan 24. De maximale gewogen scores zijn respectievelijk gelijk aan 80, 68 en 83. Het heeft de voorkeur om bij analyses *binnen* een onderwijstype (of toetsversie) gebruik te maken van de ruwe scores. Analyses over onderwijstypen heen vinden bij voorkeur plaats op basis van de bankscores die in het databestand zijn opgenomen.

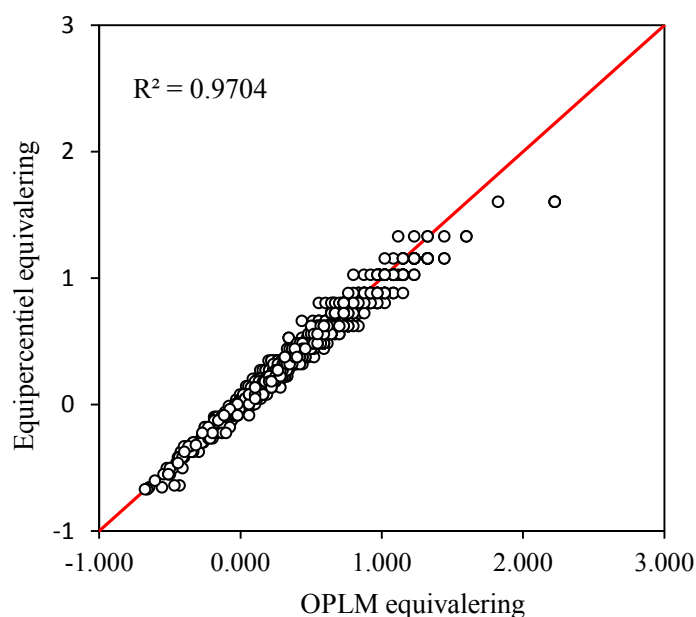
7.7 Koppeling met het Cito volgsysteem VO

Het Cito Volgsysteem VO bevat voor het derde leerjaar drie verschillende toetsen Engels Leesvaardigheid. De toetsen zijn bedoeld voor verschillende onderwijstypen en bevatten allen 40 items. De eerste toets is

bedoeld voor leerlingen in de basisberoepsgerichte leerweg van het VMBO. In het databestand is aan deze toets nummer 10 toegekend. De tweede toets (**ENG_VERS** = 11) is bedoeld voor leerlingen in de kaderberoepsgerichte, en gemengde en theoretische leerweg van het VMBO. De derde toets wordt afgenomen bij HAVO en VWO leerlingen (**ENG_VERS** = 12). De items in de toetsen Engels Leesvaardigheid liggen niet op dezelfde meetschaal als de items van COOL⁵⁻¹⁸. We passen daarom wederom een equivaleringsprocedure toe om het Cito Volgsysteem VO te relateren aan COOL⁵⁻¹⁸. In het onderhavige geval is verondersteld dat de toetsen Engels Leesvaardigheid van COOL⁵⁻¹⁸ en het Cito Volgsysteem VO gemaakt zijn door statistisch equivalente groepen. Vervolgens is het OPLM geschat met behulp van *marginal maximum likelihood*(MML). Deze procedure is eerder ook toegepast bij begrijpend lezen, taalverzorging en wiskunde (zie bijvoorbeeld paragraaf 4.8).

In totaal hebben 7 scholen van 1228 leerlingen toetsgegevens ingestuurd. Sommige leerlingen bleken geen toets Engels te hebben gemaakt. Uiteindelijk konden we per toetsversie beschikken over de volgende hoeveelheid gegevens: (ENG_VERS 10) 17 leerlingen, (ENG_VERS 11) 301 leerlingen, en (ENG_VERS 12) 910 leerlingen. Met 17 waarnemingen op ENG_VERS 10 is er geen stabiele basis om de toets te relateren aan COOL⁵⁻¹⁸. Er is dan ook besloten om daarvan af te zien. Bij de twee andere toetsversies hadden we wel voldoende gegevens en is het OPLM geschat met behulp van MML. Uit de statistische toetsen bleek dat het OPLM een adequate beschrijving geeft van de prestaties van de leerlingen: $R^2 = 243.24$, $df = 312$, $p = .99$. Het lijkt dan ook zonder meer mogelijk om de scores die de leerlingen behaald hebben op de toetsen Engels Leesvaardigheid van het Cito Volgsysteem VO met behulp van dit model te vertalen naar de meetschaal van COOL⁵⁻¹⁸.

De OPLM equivalering is gecontroleerd op basis van een (klassieke) equipercentieequivalering zonder smoothing (zie ook paragraaf 4.8). Bij de equipercentieequivalering is geen onderscheid gemaakt in onderwijstypen. Dit betekent dat het relatieve niveau van leerlingen is bepaald in de totale groep met leerlingen die toetsversie 11 of 12 hebben gemaakt. Figuur 7.4 laat zien hoe de equipercentieequivalering zich verhoudt tot de equivalering met het OPLM. We zien dat er sprake is van een sterke lineaire samenhang, $r = .97$. Voor het bepalen van de geëquivalenteerde scores van de leerlingen maakt het dus weinig uit welke aanpak we kiezen. Wij hebben ervoor gekozen om gebruik te maken van de scores die voortvloeien uit de OPLM equivalering.



Figuur 7.4. Equipercentieequivalering vergeleken met een OPLM equivalering

Net als bij eerdere toetsversies bevat het databestand voor toetsversies 11 en 12 de ruwe scores (**ENG_RSCO**, max. versie 11 = 40 / max. versie 12 = 40), de gewogen scores (**ENG_WSCO**, max. versie 11 = 154 / max. versie 12 = 142), de vaardigheidsscores (**ENG_THET**) en de bankscores (**ENG_BANK3**, max. = 100). De ruwe en gewogen scores hebben betrekking op toetsversies 11 en 12 en kunnen niet vergeleken worden met de ruwe en gewogen scores van de andere toetsversies. Op basis van de bank- en vaardigheidsscores kan die vergelijking wel gemaakt worden. Het heeft de voorkeur om bij analyses *binnen* een toetsversie gebruik te maken van de ruwe scores. Analyses over toetsversies of onderwijstypen heen vinden bij voorkeur plaats op basis van de bankscores die in het databestand zijn opgenomen.

8 BURGERSCHAPSCOMPETENTIES

8.1 Inleiding

Aan deelnemende scholen is gedurende het wervingsproces de keuze voor de afname van het vierde toetsinstrument voorgelegd (zie ook de eerste alinea van paragraaf 2.2). De school met een reguliere deelname kon kiezen tussen afname van de burgerschapsvragenlijst of afname van de toets Engels (of beide). Het burgerschapsinstrument is in deze ronde door dit keuzeprocess bij ruim een derde van het aantal COOL⁵⁻¹⁸ leerlingen afgenomen.

Het meetinstrument voor burgerschapscompetenties voor jongeren van 11 tot 16 jaar is ontwikkeld door het Instituut voor de Lerarenopleiding en het SCO-Kohnstamm Instituut van de Universiteit van Amsterdam. Naast de in COOL⁵⁻¹⁸ gebruikte vragenlijst voor leerlingen, is er ook een vragenlijst voor docenten beschikbaar. De vragenlijst voor docenten is COOL⁵⁻¹⁸ VO-3 niet gebruikt. Voor de achtergronden van het instrument in zijn geheel verwijzen we naar Ten Dam, Geijssel, Reumerman en Ledoux (2011). Gebruikers van de in dit hoofdstuk beschreven (sub)schalen dienen naar deze publicatie te verwijzen. We vermelden hier tevens dat dezelfde vragenlijst in het kader van COOL⁵⁻¹⁸ is afgenomen bij leerlingen in groep 8 van het basisonderwijs. Zie hiervoor het overeenkomstige hoofdstuk in het technisch rapport met betrekking tot het bestand van het primair onderwijs (Driessen, Mulder, & Roeleveld, 2012).

In het meetinstrument is het begrip ‘competentie’ uiteengelegd in vier componenten: kennis, reflectie, vaardigheden en attituden. Het begrip burgerschap is geoperationaliseerd aan de hand van vier centrale sociale taken: democratisch handelen, maatschappelijke verantwoordelijkheid, omgaan met conflicten, en omgaan met verschillen. Elk van deze taken is nader gedefinieerd, in algemene zin én voor de vier componenten afzonderlijk. Tabel 8.1 bevat alle definities. Naast de in deze tabel geïmpliceerde schalen bevat het instrument vijf items die het concept ‘sociale wenselijkheid’ in kaart brengen. Het kan in bepaalde analyses raadzaam zijn om voor de score op sociale wenselijkheid te controleren.

Voor elke cel in Tabel 8.1 zijn vragen voor de leerlingen geconstrueerd. Voor de componenten reflectie, vaardigheden en attituden is gebruik gemaakt van Likert-schalen (vierpuntsschaal). Voor de kenniscomponent zijn multiple-choice items gemaakt (met drie antwoordmogelijkheden, waarvan één het juiste antwoord is). De items zijn geformuleerd op het taalniveau van leerlingen uit groepen 7 en 8 van het basisonderwijs.

Op basis van exploratieve en confirmatorische factoranalyses op verschillende versies van het instrument (zie Ten Dam et al., 2011) is een vragenlijst ontstaan die bestaat uit 94 items.

Tabel 8.1*Conceptuele definities burgerschapscompetenties per component per sociale taak*

	Kennis <i>Een jongere met deze kennis ..</i>	Reflectie <i>Een jongere met deze reflectie ..</i>	Vaardigheden <i>Een jongere met deze vaardigheid ..</i>	Attituden <i>Een jongere met deze attitude ..</i>
Democratisch handelen <i>Het aanvaarden van en bijdragen aan een democratische samenleving</i>	Weet wat democratische principes zijn en wat het handelen volgens die principes inhoudt.	Denkt na over (on)democratische kwesties en kwesties van (on)macht en (on)gelijke rechten.	Kan het eigen standpunt naar voren brengen en luisteren naar de standpunten van anderen.	Wil ieders stem horen, dialoog aangaan en een actieve, kritische bijdrage leveren.
Maatschappelijke verantwoordelijkheid <i>Medeverantwoordelijkheid nemen voor de leefgemeenschappen waartoe men behoort</i>	Kent sociale regels (wettelijke of ongeschreven regels voor het sociaal verkeer).	Denkt na over belangentegenstellingen, sociale cohesie, sociale processen (in- en uitsluiting) en eigen bijdrage aan sociale rechtvaardigheid.	Kan zich sociaal rechtvaardig opstellen	Wil zich sociaal rechtvaardig opstellen (niemand uitsluiten), is bereid tot zorg en hulp, wil de ander en het milieu niet schaden door eigen gedrag.
Omgaan met conflicten <i>Betreft (lichte) conflictsituaties of belangentegenstellingen waarbij de jongere zelf 'partij' is</i>	Kent manieren om conflicten op te lossen zoals zoeken naar win-win oplossingen, hulp van anderen inroepen, ongelijk bekennen, escalatie voorkomen.	Denkt na over hoe het conflict heeft kunnen ontstaan, over eigen en andermans rol daarin en over mogelijkheden om conflicten te voorkomen of op te lossen.	Kan naar de ander luisteren, zich in de ander verplaatsen en win-win oplossingen zoeken.	Wil conflicten onderzoeken, is bereid het standpunt van de ander serieus te nemen en samen naar een acceptabele oplossing zoeken.
Omgaan met verschillen <i>Betreft sociale, culturele, religieuze en uiterlijke verschillen</i>	Kent verschillen van culturele aard, heeft kennis van gedragsregels in verschillende sociale situaties, weet wanneer er sprake is van vooroordeel en discriminatie.	Denkt na over aard en gevolgen van verschillen tussen mensen, culturele achtergronden van gedrag en processen van in- en uitsluiting.	Kan zich bewegen in onbekende sociale situaties, zich aanpassen aan andermans wensen of gewoonten.	Wil andermans opvattingen en leefstijl leren kennen, staat positief tegenover verschillen.

27 kennis-items: Deze items gaan over weten, begrijpen, inzicht hebben in wat je het beste kunt doen in een specifieke situatie. Bij de kennis-items moesten de leerlingen bij elke vraag het beste antwoordalternatief aankruisen, bijvoorbeeld: 'Alle kinderen hebben recht ... a) op zakgeld, b) om te kiezen bij wie je woont, c) op onderwijs'. [In dit geval is optie c het goede antwoord.]

28 reflectie-items: Deze items gaan over reflecteren (nadenken over). De voorgelegde vraag is: 'Hoe vaak denk jij na over...?', bijvoorbeeld '...of er naar leerlingen wordt geluisterd op jouw school?' De antwoordopties zijn: (bijna) nooit, heel af en toe, vrij vaak, vaak.

15 vaardigheden-items: Dit zijn 'self-efficacy' items: de leerling moest aangeven hoe goed hij/zij zichzelf ergens in vindt. De voorgelegde vraag is 'Hoe goed ben jij in...', bijvoorbeeld '...bij een ruzie een oplossing vinden waarmee iedereen tevreden is?'. De antwoordopties zijn: helemaal niet goed, niet zo goed, best wel goed, heel goed.

24 attitude-items: Deze items gaan over iets vinden (menen), willen, tot iets bereid zijn. De voorgelegde vraag is 'Hoe goed past een uitspraak bij jou?'. Een voorbeelduitspraak is 'Ik vind het leuk om iets te weten van verschillende soorten geloof'. De antwoordopties zijn: past helemaal niet bij mij, past niet erg bij mij, past best wel wat bij mij, past helemaal bij mij.

De items zijn overwegend positief geformuleerd. Om antwoordtendentie tegen te gaan, zijn bij de component attitude drie extra negatieve items opgenomen. Bij de component reflectie zijn om dezelfde reden vier dummy-items opgenomen. Dit zijn items die inhoudelijk niet tot het te meten domein horen en die daardoor de respondent ‘dwingen’ even opnieuw na te denken over het te geven antwoord. Verder bevat de vragenlijst bij de component attitude vijf items voor het meten van sociale wenselijkheid. Deze in totaal twaalf extra items maken geen deel uit van de bedoelde schalen en zijn dus bij de schaalconstructie genegeerd. Van de vijf sociale wenselijkheid items is een aparte schaal gemaakt.

In totaal zijn van 8188 leerlingen bruikbare gegevens verkregen. Het bestand bevat een dummy variabele **DUM_BS**, die aangeeft of de leerling wel of niet op de burgerschapsschalen vertegenwoordigd is. Omwille van een optimale vergelijkbaarheid binnen COOL⁵⁻¹⁸ zijn precies dezelfde schalen geconstrueerd als in het basisonderwijs, en wel op precies dezelfde wijze. Dat laatste slaat vooral op het aantal items per sub-schaal dat minimaal beantwoord moest zijn, voordat als schaalscore het itemgemiddelde is berekend. Bij de kennisschalen zijn, zoals gebruikelijk, ontbrekende antwoorden fout gerekend. Op de hoofdschalen voor de drie overige componenten (attitude, vaardigheid, reflectie) is het minimum aantal items dat beantwoord moest zijn, gesteld op 3, en op de diverse sub-schalen op 2. In afwijking van het hierboven gestelde hebben we voor de kennisschalen alleen een schaalscore berekend indien minstens de helft van de betreffende items was beantwoord.

8.2 Schalen

Op de items zijn exploratieve en bevestigende factoranalyses, alsmede betrouwbaarheidsanalyses uitgevoerd. Zie hiervoor de publicatie van Ten Dam et al. (2011). Er zijn in totaal 24 schalen en sub-schalen gevormd, inclusief de schaal voor sociale wenselijkheid. In principe komen deze schalen overeen met het in Tabel 8.1 gegeven schema. In de eerste plaats zijn vier hoofdschalen gemaakt, die overeenkomen met de kolommen van de tabel. Dit zijn de overall schalen kennis, attitude, reflectie en vaardigheid, met als namen **BS_KTOT**, **BS_ATOT**, **BS_RTOT** en **BS_VTOT**.

Verder zijn er sub-schalen gemaakt, die in principe overeen komen met de 16 combinaties van componenten met inhoudelijke domeinen. Hierop zijn de volgende drie uitzonderingen. Ten eerste zijn er, naast de sub-schaal voor democratisch handelen binnen de attitudinale component, twee deel-sub-schalen, waarvan de eerste betrekking heeft op ‘ieders stem willen horen’, en de tweede op een ‘kritische bijdrage willen leveren’. Ten tweede geldt iets soortgelijks voor de sub-schaal voor democratisch handelen binnen de vaardigheidscomponent. Naast de betreffende sub-schaal zijn er twee deel-sub-schalen, waarvan de eerste betrekking heeft op het ‘opkomen voor de eigen mening’, en de tweede op het ‘luisteren naar de mening van anderen’. Uit de factoranalyses op de betreffende itemsets kwamen in beide gevallen twee interpreteerbare factoren naar voren. Ten derde zijn binnen het vaardigheidsdomein de twee sub-schalen voor maatschappelijk verantwoord handelen en omgaan met conflicten samengenomen. Deze sub-schalen, bestaande uit respectievelijk twee en drie items bleken juist geen duidelijk onderscheidbare factoren op te leveren. Tabel 8.2 geeft een overzicht van de in totaal 24 (sub)schalen.

Tabel 8.2*Overzicht schalen burgerschap*

naam	#items	α	N valide	M	SD
BS_ATOT	24	.89	4400	2.83	0.39
BS_VTOT	15	.83	4384	3.02	0.34
BS_RTOT	28	.94	4393	2.16	0.52
BS_KTOT	27	.85	4363	0.83	0.16
BS_ADH	6	.73	4396	2.86	0.45
BS_ADH_1	3	.76	4395	3.21	0.48
BS_ADH_2	3	.70	4393	2.49	0.63
BS_AMV	6	.67	4395	2.94	0.45
BS_ACO	6	.80	4391	2.80	0.50
BS_AVE	6	.87	4388	2.72	0.62
BS_VDH	6	.65	4383	3.04	0.39
BS_VDH_1	3	.77	4381	3.14	0.55
BS_VDH_2	3	.66	4378	2.93	0.49
BS_VMVCO	5	.76	4379	2.96	0.45
BS_VVE	4	.69	4375	3.09	0.44
BS_RDH	6	.80	4391	2.13	0.58
BS_RMV	6	.85	4387	2.00	0.65
BS_RCO	8	.91	4383	2.52	0.66
BS_RVE	8	.87	4382	1.92	0.62
BS_KDH	8	.65	4375	0.88	0.17
BS_KMV	6	.57	4363	0.83	0.21
BS_KCO	7	.67	4359	0.73	0.24
BS_KVE	6	.63	4332	0.89	0.19
BS_SW	5	.61	4270	2.58	0.43

De eerste kolom van deze tabel bevat de variabelenamen, die weer maximaal acht posities innemen en telkens beginnen met **BS_**. De volgende letter in de naam van de eerste vier hoofdschalen in het bovenste blok verwijst naar de ‘component’, waarbij **A** staat voor ‘attitude’, **V** voor ‘vaardigheid’, **R** voor ‘reflectie’, en **K** voor ‘kennis’. De toevoeging **TOT** geeft aan dat het de totale schaal is. De volgende vier blokken in Tabel 8.2 bevatten de sub-schalen voor respectievelijk de vier componenten. De lettercombinaties geven de specifieke sub-schaal aan. **ADH** is dus de sub-schaal voor democratisch handelen binnen de attitudinale component. De twee deel-sub-schalen worden aangeduid als **ADH_1** en **ADH_2**. Hetzelfde geldt voor **VDH_1** en **VDH_2**, de twee deel-sub-schalen voor democratisch handelen binnen de vaardigheidscomponent. **VMVCO** geeft aan dat het de combinatie van maatschappelijk verantwoord handelen en omgaan met conflicten binnen de vaardigheidscomponent betreft. In de onderste regel staat de sociale wenselijkheidschaal (**BS_SW**).

De tweede kolom van Tabel 8.2 geeft het aantal items van elke (sub)schaal, de derde kolom de betrouwbaarheid (α), de vierde kolom het aantal leerlingen dat een schaalscore toegekend heeft gekregen, de vijfde kolom het gemiddelde, en de zesde kolom de bijbehorende standaard deviatie. Voor de kennisschalen geldt hierbij dat de betrouwbaarheid is bepaald op basis van de gegevens van leerlingen die ten minste de helft van de items hebben beantwoord.

9 LEERLINGVRAGENLIJST

9.1 Inleiding

De in de leerlingvragenlijst opgenomen variabelen dienen ter operationalisatie van de in het onderzoeksvoorstel opgenomen concepten. De leerlingen op scholen die aan de gecombineerde Inventaar en COOL⁵⁻¹⁸ afname hebben meegedaan hebben een aantal vragen gekregen die specifiek voor Inventaar waren. In volgorde van verschijnen in de vragenlijst betreft het de volgende concepten: welbevinden met docenten; lestijd; een verkorte versie van de ‘big five’ persoonlijkheidsfactoren van de ‘Five Factor Personality Inventory’ (FFPI); welbevinden met klasgenoten; de acht motivationele schalen, die samen de Inventory of School Motivation (ISM) vormen; self-efficacy; taakoriëntatie.

De uitwerking van deze concepten komt respectievelijk in paragrafen 9.3.1 tot en met 9.3.7 ter sprake. In deze paragrafen worden ook de bronnen vermeld, waarnaar bij gebruik van de betreffende concepten verwezen dient te worden. Daarvoor worden in paragraaf 9.2 enkele procedures toegelicht. Naast de in het onderzoeksvoorstel opgenomen concepten heeft het GION ook zelf een aantal concepten in de vragenlijst opgenomen. De betreffende variabelen zijn niet in het COOL⁵⁻¹⁸ bestand opgenomen en komen derhalve ook niet in dit rapport aan de orde.

9.2 Procedures

De door de scholen naar Cito geretourneerde vragenlijsten zijn daar optisch ingelezen, volgens de wensen van het GION. Deze wensen betroffen het omgaan met dubbele antwoorden. Hoewel in de invulinstructie expliciet staat vermeld dat er per vraag slechts één antwoord mocht worden gegeven, is het de ervaring van het GION dat ook dan sommige respondenten bij sommige vragen meer dan één antwoord aankruisen (of omcirkelen). In verreweg de meeste van deze gevallen gaat het daarbij om twee aanliggende antwoorden op items die een Likert-achtig karakter hebben. De enig zinvolle interpretatie van dergelijke antwoorden is dat de respondent heeft willen aangeven dat zijn/haar antwoord “er tussen in” ligt. Aangezien dergelijke antwoordpatronen zinvol verwerkt kunnen worden, zijn ze ook met unieke codes ingevoerd. Dit wijkt af van de standaard benadering bij optisch inlezen – en veelal ook bij handmatige invoer – waarbij dubbele antwoorden hetzij direct gelijk worden gesteld aan ‘geen antwoord’ of ‘fout antwoord’, hetzij één algemene code voor ‘dubbel antwoord’ krijgen toegekend. Na het inlezen zijn de data als een EXCEL-bestand aan het GION overgedragen. De verdere werkzaamheden zijn door het GION verricht. Deze werkzaamheden betreffen het omzetten naar SPSS, het toekennen van variabelen namen en het ‘labelen’ van de coderingen, het hercoderen van met name de coderingen die een dubbel antwoord representeren, het definiëren van ontbrekende data coderingen, en het aanmaken van schaalscores.

9.2.1 Invulinstructie

De vragenlijst begon met de volgende invulinstructie:

“Deze vragenlijst is onderdeel van het COOL5-18 onderzoek waarvoor je vermoedelijk ook al een aantal toetsen hebt gemaakt. Alle vragen in deze lijst zijn vertrouwelijk. Dat betekent dat de onderzoekers er zeer zorgvuldig mee omgaan en dat niemand anders de antwoorden te zien krijgt. Ook de school of je docenten krijgen je antwoorden niet te zien.

*Neem rustig de tijd voor het lezen en beantwoorden van de vragen. Probeer elke vraag te beantwoorden, ook wanneer je het soms moeilijk vindt. Het gaat steeds om jouw mening of om dingen die je zelf wilt of doet. Je moet bij elke vraag één antwoord geven. Dat doe je door in het vakje van dat antwoord met een zwarte potlood een kruis te zetten. **Kruis steeds het hokje aan dat het beste bij jouw mening past.** Als je je vergist hebt, gum je het foute antwoord uit en kruis je een ander hokje aan.”*

9.2.2 Ontbrekende waarden

Welke codering voor ontbrekende waarden is gebruikt, hangt af van het aantal antwoordmogelijkheden op het betreffende item. Als dat hoogstens 7 bedraagt, is de codering 9 gebruikt, bij een hoger aantal antwoordmogelijkheden is de codering 99 gebruikt. De labeling van de betreffende codering is enigszins afhankelijk van de vraag: ‘ontbrekend’ of ‘geen antwoord’. In een aantal gevallen horen de items binnen een concept zodanig bij elkaar dat het zinvol is om onderscheid te maken tussen leerlingen die één of enkele items niet hebben beantwoord, en leerlingen die geen enkel item hebben beantwoord. In deze gevallen geldt: 9 of 99 = hele vraag niet beantwoord, 8 of 98 = specifiek item niet beantwoord, 7 of 97 = onbruikbare score.

9.2.3 Dubbele coderingen

Zoals al aangegeven zijn dubbele antwoorden specifiek, ofwel ‘als zodanig’, ingelezen, en kan er onderscheid worden gemaakt tussen ‘zinvol dubbel’ en ‘anders dubbel’. De ‘zinvol dubbele’ antwoorden betreffen het zijn aangekruist van twee naast elkaar liggende antwoordcategorieën in geval van Likert-achtige items. De coderingen die dergelijke antwoordpatronen aangeven zijn gehercodeerd naar de waarde die de respondent redelijkerwijs bedoeld heeft. Indien bijvoorbeeld de 4 en 5 beide zijn aangekruist, is dat ingelezen als 45, en gehercodeerd naar 4.5. In de overige gevallen, indien twee niet aanliggende antwoordcategorieën zijn aangekruist, of wanneer er meer dan twee categorieën zijn aangekruist, is de betreffende dubbelcodering gehercodeerd naar 8 of 98, welke codes gelabeld zijn als ‘overig’ en als ‘ontbrekend’ zijn gedeclareerd.

9.2.4 Constructie schaalscores

Voor de diverse in de vragenlijst opgenomen itemsets die elk één of meerdere onderliggende concepten meten, is nagegaan hoeveel items per beoogde (sub)schaal zijn beantwoord. Tevens is per (sub)schaal een betrouwbaarheidsanalyse uitgevoerd. Een beslispunt bij het aanmaken van de schaalscores, is het maximale aantal items dat eventueel niet beantwoord mag zijn. In de meest strikte opvatting moeten alle items van een bepaalde schaal zijn beantwoord, voordat als schaalscore het gemiddelde over de items mag worden berekend. Dit leidt ertoe dat relatief veel leerlingen een ontbrekende schaalscore moeten krijgen toegekend. In de praktijk is het daarom gebruikelijk om soepeler te zijn, door bijvoorbeeld één of enkele ontbrekende itemscores toe te staan. Wij hebben getracht een redelijk evenwicht te vinden tussen de volgende twee desiderata: enerzijds optimale nauwkeurigheid, anderzijds een schaalscore toekennen aan zo veel mogelijk leerlingen. We hebben in de meeste gevallen één ontbrekende waarde toegestaan. Alleen bij de verkorte

versie van FFPI is hiervan afgeweken en is aangesloten bij de voor dit instrument gangbare praktijk. Het spreekt voor zich dat bepaalde negatief geformuleerde items eerst gehercodeerd (gespiegeld) zijn, voordat het itemgemiddelde is berekend.

9.3 Systematische beschrijving

De vragenlijst was ontworpen om optisch te worden ingelezen. De antwoordmogelijkheden waren vierkante hokjes die moesten worden aangekruist. Er werd dus geen gebruik gemaakt van cijfers. In de volgende sub-paragrafen maken we echter wel gebruik van cijfers om de antwoordmogelijkheden mee aan te duiden. Dit zijn tevens de coderingen die in het bestand zijn gebruikt. De antwoordmogelijkheden zelf zijn doorgaans gebruikt als 'value labels'. Het leerlingvragenlijstdeel van het SPSS bestand bevat de gegevens van 14.394 leerlingen, die tenminste één vraag van de vragenlijst hebben beantwoord. In het totale SPSS bestand geeft de variabele **DUM_LV** aan of dit al dan niet het geval is. De in de rest van dit hoofdstuk vermelde percentages hebben in principe betrekking op deze 14.394 cases.

9.3.1 Welbevinden met docenten

De schaal 'welbevinden met docenten' bestaat uit zeven items, die afkomstig zijn van Peetsma, Wagenaar, en De Kat (2001). Bij gebruik van deze items en/of de erop gebaseerde schaal dient naar deze publicatie te worden verwezen. De items betreffen de mening van de leerlingen over hun docenten en werden als volgt ingeleid: *"Wat vind je van de docenten (leraren en leraressen) op je school?"*. Daaronder stonden de volgende items, die hier – maar niet in de vragenlijst – worden voorafgegaan door de toegekende naam:

LV_WDO1	De docenten weten meestal wel hoe ik me voel
LV_WDO2	Ik kan met de docenten over problemen praten
LV_WDO2	Als ik me ongelukkig voel, kan ik daar met de docenten over praten
LV_WDO4	Ik voel me bij de docenten op mijn gemak
LV_WDO5	De docenten begrijpen mij
LV_WDO6	Ik heb een goed contact met de docenten
LV_WDO7	Ik zou liever andere docenten hebben

De antwoordmogelijkheden waren als volgt: 1 = klopt helemaal niet, 2 = klopt niet, 3 = klopt soms wel, soms niet, 4 = klopt, 5 = klopt precies. In de vragenlijst ontbraken de cijfers, en stonden de antwoordmogelijkheden boven de vijf kolommen met de antwoordhokjes. De cijfers geven de coderingen in het bestand aan. Dit 'klopt' regime van antwoorden is bij veel van de andere itemsets in de vragenlijst ook gebruikt. Van de zeven uitspraken zijn er zes positief geformuleerd, waarbij een hoge score staat voor een hoge mate van tevredenheid met de docenten, en is één item, het laatste, negatief geformuleerd, hetgeen inhoudt dat een hoge score staat voor ontevredenheid met de docenten.

Uit de telling van het aantal valide antwoorden binnen leerlingen over de zeven items, blijkt dat 14.032 leerlingen (97,5%) alle items hebben beantwoord, dat 286 leerlingen (0,2%) één item hebben overgeslagen en dat 36 leerlingen (< 0,1%) geen enkel item hebben beantwoord. De overige 40 leerlingen (< 0,1%) hebben meerdere items overgeslagen. Na spiegeling van het negatief geformuleerde zevende item bedraagt de betrouwbaarheid .84. Er zijn geen items die de betrouwbaarheid verlagen. Aan het bestand is een nieuwe

variabele **LV_WDOC** toegevoegd, waarvan de score is berekend door het gemiddelde te nemen van de zeven items, waarbij de score op één item mocht ontbreken. Op deze variabele hebben 14.318 leerlingen (99,5%) een score. Het gemiddelde bedraagt 3.11 ($SD = 0.63$).

9.3.2 Lestijd

In de vragenlijst waren twee vragen gewijd aan het onderwerp lestijd. Deze vragen werden niet apart ingeleid. De eerste vraag (**LV_LES1**) luidde: *“Hoeveel lesuren per week moet je op school zijn?”* Voor het antwoord waren twee antwoordhokjes gereserveerd. De vraag is niet beantwoord door 758 leerlingen (0,5%). De wel gegeven antwoorden variëren van 0 tot en met 99. Dat betekent dat er ongeloofwaardig lage en hoge antwoorden voorkomen. Het gemiddelde is 32 uur ($SD = 4.2$).

De tweede vraag betrof een verdeling van de lesuren over verschillende activiteiten, en luidde: *“In een gemiddelde week: Hoeveel lesuren ...”*. Daaronder stonden de volgende vijf items:

- LV_LES2A** krijg je uitleg en instructie (les) van een docent
- LV_LES2B** moet je opdrachten en werkstukken maken onder begeleiding van een docent
- LV_LES2C** moet je ‘huiswerk’ maken of proefwerken leren onder leiding van een docent
- LV_LES2D** mag je (in de klas) zelf weten wat voor schoolwerk je doet
- LV_LES2E** vallen uit (ziekte/afwezigheid docent, andere redenen)

Er waren de volgende antwoordmogelijkheden: 1 = geen enkele, 2 = 1-2 uur, 3 = 3-5 uur, 4 = 6-10 uur, 5 = 11-15 uur, 6 = 16-20 uur, 7 = 21-25 uur, 8 = 26-30 uur. In totaal hebben 179 leerlingen (0,1%) voor geen enkele van deze vijf onderdelen iets ingevuld, met andere woorden zij hebben deze vraag in het geheel niet beantwoord. Deze leerlingen hebben op elk van de vijf items de ontbrekende waarde code 99 toegekend gekregen, die gelabeld is als ‘hele vraag niet beantwoord’. Daarnaast komen op elk onderdeel specifieke ontbrekende waarden voor, respectievelijk 63, 82, 54, 49, 48. Hiervoor is de ontbrekende waarde code 98 gebruikt, die gelabeld is als ‘item niet beantwoord’. De meest voor de hand liggende interpretatie van de specifieke ontbrekende waarden is dat ze inhoudelijk hetzelfde betekenen als het antwoord ‘geen enkele’. Gebruikers van het bestand kunnen desgewenst zelf een dergelijke hercodering aanbrengen. Voor dubbele antwoorden, anders dan die als 1.5, 2.5, etc. konden worden gehercodeerd, is de ontbrekende waarde code 97 toegekend, die gelabeld is al ‘overig/onbekend’.

9.3.3 Persoonlijkheid

Het persoonlijkheidsprofiel van de leerlingen is vastgesteld met een subset van 52 items afkomstig uit de ‘Five-Factor Personality Inventory’ (FFPI; Hendriks, Hofstee, & De Raad, 1999, 2011), waarmee scores worden verkregen op de vijf factoren Extraversie, Mildheid, Ordelijkheid, Emotionele Stabiliteit en Intellectuele Autonomie. Extraversie beschrijft iemands mate van sociale deelname, energieniveau en expressiviteit. Mildheid beschrijft hoe iemand met anderen omgaat. Ordelijkheid beschrijft hoe iemand met taken, spullen en zijn of haar omgeving omgaat. Emotionele Stabiliteit beschrijft hoe bestendig iemand is tegen emotionele en fysieke prikkels, zoals indrukken, drukte en geluiden. Intellectuele Autonomie beschrijft iemands mate van onafhankelijkheid in denken en beslissen. De FFPI is oorspronkelijk voor

volwassenen ontwikkeld, maar bleek met enkele kleine aanpassingen tevens geschikt voor leerlingen vanaf ongeveer 11 jaar (Hendriks et al., 2011; Hendriks, Kuiper, Offringa & van der Werf, 2008).

De volledige FFPI bestaat uit 100 korte gedragsbeschrijvende zinnen. Per factor zijn er 20 items, waarvan er 10 de positieve ‘pool’ betreffen, en 10 de negatieve. De subset van items uit FFPI die in de leerlingvragenlijst voor VO-3 meting in 2014 zijn opgenomen betreffen 10 items per factor. Deze items zijn geselecteerd op basis van de factorladingen in een voorgaande afname van COOL5-18 in VO-3 in 2011. De 50 items zijn aangevuld met 2 aanvullende items voor de factor Intellectuele Autonomie; de factor die over het algemeen de laagste interne consistentie heeft. De lengte van de FFPI is de voornaamste reden geweest voor het selecteren van een subset van items. Dezelfde subset van items is ook afgenomen in de HAVO-5 afname in 2013 en de VWO-6 afname in 2014.

In de vragenlijst waren de items verspreid over twee plaatsen, met elk 26 items. De reden hiervoor was het aanbrengen van enige afwisseling in de vragenlijst wat betreft het soort vragen dat moest worden beantwoord. De eerste keer (vraag 12) werden de persoonlijkheidsvragen-items als volgt ingeleid: *‘De volgende uitspraken gaan over hoe je zelf vindt dat je bent. Het vraagteken betekent ‘geen nee en geen ja’. Denk niet te lang na: je eerste gedachte is vaak het best. Wanneer je een uitspraak (echt) niet snapt, mag je het hokje ‘snap ik niet’ aankruisen.’* De tweede keer (vraag 19) was de inleiding hetzelfde, met als enige verschil dat in de eerste zin het woordje ‘weer’ was toegevoegd: *‘De volgende uitspraken gaan weer ...’*.

De antwoordmogelijkheden stonden boven de kolommen met hokjes en luiden: NEE! nee ? ja JA! en ‘snap ik niet’. De antwoorden zijn ingelezen met codes 1 tot en met 5 voor de vijf inhoudelijke antwoorden en code 6 voor ‘snap ik niet’. Bij twee aanliggende antwoorden werd in dit geval de meest extreme van de twee codes genomen, dus bij 12 werd dit 1, bij 23 werd dit 2, bij 34 werd dit 4 en bij 45 werd dit 5. Indien zowel een itemantwoord was gegeven maar ook ‘snap ik niet’ was aangekruist, is gekozen voor het ‘snap ik niet’-antwoord. De andere mogelijkheden zijn als ontbrekend behandeld. Enkele voorbeelditems zijn:

- Ik knoop gesprekken aan (Extraversie)
- Ik probeer ruzie te voorkomen (Mildheid)
- Ik werk graag volgens schema (Ordelijkheid)
- Ik kan problemen van me afzetten (Emotionele Stabiliteit)
- Ik weet wat ik wil (Intellectuele Autonomie)

Doordat slechts een subset van de items is afgenomen kunnen factorscores niet op dezelfde manier bepaald worden als in de vorige afnamen van COOL⁵⁻¹⁸. Om te bepalen hoe op basis van de subset van 52 items het beste factorscores bepaald kon worden is aanvullend onderzoek gedaan. We verwijzen hiervoor naar bijlage 4 van het technisch rapport over de havo-5 afname (Keuning, Zijlsing, Naayer, & Timmermans, 2015).

Aan het bestand zijn nieuwe variabelen **LV_BF_FACT1** tot en met **LV_BF_FACT5** (Tabel 4.5) toegevoegd waarvan de score is berekend door het bepalen van een somscore (na spiegeling van de antwoorden op de items van de negatieve ‘pool’), waarbij maximaal 2 itemscores per factor mocht ontbreken (maximaal 1 per factorpool). Bij de laatste factor (Intellectuele Autonomie) mochten maximaal drie itemscores ontbreken (maximaal 2 per factorpool). Indien leerlingen ontbrekende waarden hadden op enkele items, maar verder voldeden aan de voorwaarde van het aantal valide observaties per factorpool is

een imputatie verricht van de ontbrekende items. In deze gevallen is het gemiddelde bepaald per factorpool en is vervolgens het dichtstbijzijnde gehele getal geïmputeerd. Daarna zijn de somscores middels een equipercentiel equivalering (zie bijvoorbeeld Kolen & Brennan, 2004) omgezet naar de meetschaal van de oorspronkelijke FFPI. Daarmee wordt het mogelijk om de factorscores van de vijf persoonlijkheidsfactoren in deze afname van COOL⁵⁻¹⁸ te vergelijken met de factorscores uit voorgaande afnamen. De vergelijking brengt echter wel enige onzekerheid met zich mee. We nemen immers slechts een subset van items af en veronderstellen vervolgens dat het relatieve niveau van leerlingen op deze subset identiek is aan het relatieve niveau dat zij zouden behalen op de volledige set. Het is onduidelijk in hoeverre deze veronderstelling juist is. Daar komt bij dat bij deze afname van COOL⁵⁻¹⁸ een gelijk gewicht aan alle items is toegekend, terwijl dat bij de oorspronkelijke FFPI niet het geval is. Voor deze VO-3 afname is een afzonderlijke equivalering uitgevoerd. De resultaten van deze equivalering komen sterk overeen met de resultaten van de havo-5 en vwo-6 afname. N.B.

De equivalering is ook de reden waarom voor de factorscores enigszins andere namen zijn gekozen dan in de eerdere VO-3-afnames (daar heten ze LV_FFPI1 t/m LV_FFPI5) – om aan te geven dat de scores op het persoonlijkheidsprofiel nog wel zijn gebaseerd op de Big Five-theorie, maar dat deze scores niet 100% vergelijkbaar zijn met de ‘echte’ FFPI-factorscores.

Van de 14.394 leerlingen hebben er 13.914 (96.7%) scores op alle vijf factoren. In totaal hebben 230 leerlingen (1.6%) op geen enkele factor een score. De overige 250 leerlingen (1.7%) hebben op minimaal één en maximaal vier factoren een score. Bij de 13.993 leerlingen met valide factorscores op alle vijf factoren is onderzoek gedaan naar de betrouwbaarheid van de scores. De interne consistentie (gestratificeerde α) was .86 voor Extraversie, .82 voor Mildheid, .85 voor Ordelijkheid, .85 voor Emotionele Stabiliteit, en .80 voor Intellectuele Autonomie. Om auteursrechtelijke reden zijn alleen de scores op de FFPI factoren (en geen itemscores) in het bestand opgenomen, met de volgende namen:

Tabel 9.1

Basisinformatie FFPI: variabele naam, factor naam, aantal, gemiddelde en standaard deviatie

Variabele	Naam	<i>N</i>	<i>M</i>	<i>SD</i>
LV_BF_FACT1	Extraversie	14040	0.86	0.83
LV_BF_FACT2	Mildheid	14125	2.00	0.97
LV_BF_FACT3	Ordelijkheid	14120	0.28	1.04
LV_BF_FACT4	Emotionele Stabiliteit	14054	1.01	0.93
LV_BF_FACT5	Intellectuele Autonomie	14119	0.68	0.79

9.3.4 Welbevinden met klasgenoten

De schaal ‘welbevinden met klasgenoten’ bestaat uit zes items, die afkomstig zijn van Peetsma, Wagenaar, en De Kat (2001). Bij gebruik van deze items en/of de erop gebaseerde schaal dient naar deze publicatie te worden verwezen. De items betreffen de mening over de klasgenoten en werden als volgt ingeleid: “*Wat vind je van je klasgenoten?*”. Daaronder stonden de volgende items, die weer volgens het ‘klopt’ regime moesten worden beantwoord:

- LV_WKL1** Ik heb veel contact met mijn klasgenoten
- LV_WKL2** Ik zou liever in een andere klas zitten
- LV_WKL3** Wij hebben een leuke klas

LV_WKL4	Ik kan goed met mijn klasgenoten overweg
LV_WKL5	Ik mijn klas voel ik me soms alleen
LV_WKL6	Ik vind het leuk om met de leerlingen uit mijn klas om te gaan

Van de zes items zijn er vier positief en twee – items 2 en 5 – negatief geformuleerd. Uit de telling van het aantal valide antwoorden binnen leerlingen over de zes items, blijkt dat 14.149 leerlingen (98,3%) alle items valide hebben beantwoord, dat 189 leerlingen (0,1%) één item hebben overgeslagen en dat 36 leerlingen (< 0,1%) geen enkel item hebben beantwoord. De overige 20 leerlingen (< 0,1%) hebben meerdere items overgeslagen. Na spiegeling van de negatief geformuleerde items bedraagt de betrouwbaarheid .83. Er zijn geen items die de betrouwbaarheid verlagen. Aan het bestand is een nieuwe variabele **LV_WKLAS** toegevoegd waarvan de score is berekend door het gemiddelde te nemen van de zes items, waarbij één itemscore mocht ontbreken. Op deze nieuwe variabele hebben 14.338 (99,6%) leerlingen een score. Het gemiddelde bedraagt 3.98 ($SD = 0.67$).

9.3.5 Motivatie

De wetenschappelijke literatuur over motivatie is zeer uitgebreid. Uit de vele instrumenten die in de afgelopen decennia zijn ontwikkeld, is in COOL⁵⁻¹⁸ gekozen voor de ‘Inventory of School Motivation (ISM)’. Dit instrument is gebaseerd op het ‘personal investment model’ van Maehr (1984) en is ontwikkeld door McInerney en collega’s (McInerney & Sinclair, 1991). Bij de ontwikkeling lag de nadruk vrij sterk op toepasbaarheid binnen diverse culturen. In COOL⁵⁻¹⁸ is gebruik gemaakt van een vrij korte versie (totaal 33 items) van Ali en McInerney (2004). In totaal bevat de ISM acht sub-schalen (of ‘eerste orde factoren’), die geacht worden vier ‘tweede orde’ schalen te vormen, die geacht worden tezamen zoiets als ‘algemene motivatie’ te meten. Aangezien van de ISM nog geen Nederlandse vertaling beschikbaar was, hebben we de items zelf vertaald. Ten opzichte van de afname in 2008 is voor één item, dat de betrouwbaarheid van de betreffende sub-schaal bleek te verlagen (zie paragraaf 10.3.6 uit het Technisch Rapport over de eerste meting), een andere vertaling gemaakt. De bijlage bij dat rapport bevat alle 33 ISM-items.

De acht sub-schalen hebben in het Engels de volgende namen: task, effort, competition, social concern, social power, praise, affiliation, token. Task en effort vormen samen ‘mastery’, competition en social power vormen samen ‘performance’ – dit zijn de twee voornaamste, ook in andere benaderingen van motivatie voorkomende ‘achievement goals’ (Elliott & McGregor, 2001). Social concern en affiliation vormen samen ‘sociale motivatie’, en praise en token vormen samen ‘extrinsieke motivatie’. De overkoepelende motivatie schaal wordt ook wel Gmot genoemd, naar analogie van G als aanduiding van ‘general intelligence’. De 33 items waren verspreid over twee plaatsen (bladzijden) in de vragenlijst en werden de eerste keer als volgt ingeleid: *“Nu volgt een aantal stellingen over je werk op school, en de opdrachten die je op school krijgt”*. De antwoordmogelijkheden volgden het ‘klopt’ regime, zij het dat de middelste drie iets anders waren gelabeld: 1 = klopt helemaal niet, 2 = klopt een beetje, 3 = klopt matig, 4 = klopt vrij goed, 5 = klopt precies.

De scores op de acht sub-schalen zijn berekend als het gemiddelde van de samenstellende items, waarbij maximaal één itemscore mocht ontbreken, tenzij de sub-schaal uit slechts drie items bestond (social power en affiliation). De scores op de vier tweede orde schalen zijn berekend als het gewogen gemiddelde van de twee deel uitmakende schalen, indien beide een valide score hadden. De score op ‘algemene motivatie’ is berekend als het gewogen gemiddelde van de vier tweede orde schalen, indien deze alle een valide score

hadden. Daarnaast is een overall schaal gemaakt volgens een soepeler criterium met betrekking tot ontbrekende waarden. Op deze schaal is een score toegekend, indien maximaal zes itemscores ontbraken. De score op deze schaal is gedefinieerd als het gemiddelde over de items met een valide score. In Tabel 9.2 zijn de voornaamste kenmerken van de sub-schalen en hogere orde schalen samengevat.

Tabel 9.2
Overzicht schalen Inventory of School Motivation

Schaal	Naam	#items	α	N	M	SD
task	LV_TSK	4	.61	14244	3.77	0.67
effort	LV_EFF	5	.77	13995	2.96	0.71
competition	LV_COMP	4	.80	14004	2.32	0.91
social concern	LV_SOCO	5	.73	13985	3.12	0.72
social power	LV_SOPO	3	.83	13902	1.99	0.93
praise	LV_PRS	5	.82	13987	2.61	0.85
affiliation	LV_AFF	3	.67	13828	3.33	0.81
token	LV_TOK	4	.72	14261	2.81	0.86
mastery	LV_MAST	9	.85	13955	3.32	0.59
performance	LV_PERF	7	.84	13886	2.18	0.80
social	LV_SOCL	8	.74	13818	3.20	0.62
extrinsic	LV_EXTR	9	.85	13982	2.70	0.77
Gmot	LV_GMOT	33	.89	13686	2.88	0.51
Gmot2	LV_GMOT2	33	.89	13979	2.88	0.51

9.3.6 Self-efficacy

Het concept ‘(perceived) self-efficacy’ is vermoedelijk geïntroduceerd door Bandura (1986). In Wikipedia wordt het begrip omschreven als ‘het vermogen en de overtuiging om adequaat en efficiënt te handelen in een gegeven situatie’. Dit vermogen en deze overtuiging zullen voor een persoon doorgaans per domein verschillen. Hier gaat het uiteraard om ‘schoolse self-efficacy’. Ruim genomen kan dit worden omschreven als de mate waarin de leerling het gevoel heeft “aan te kunnen” wat er op school van hem/haar wordt gevraagd, of, met andere woorden, een goede inschatting van en vertrouwen in het eigen kunnen heeft. De gebruikte itemset is in COOL⁵⁻¹⁸ ingebracht door het Kohnstamm Instituut en maakt ook deel uit van de vragenlijst die in primair onderwijs is afgenomen. Het betreft een schaal van de PALS (Midgley et al., 2000). Bij gebruik van deze items en/of de schaal dient naar deze publicatie te worden verwezen. In de vragenlijst was geen aparte inleiding boven deze items, waarvan de antwoordmogelijkheden weer conform het eerdere ‘klopt’ regime waren. Voorgelegd werden de volgende uitspraken:

- LV_SEF1** Ik weet zeker dat dit jaar alles op school wel zal lukken
- LV_SEF2** Ik kan op school zelfs de moeilijkste opdrachten maken als ik mijn best doe
- LV_SEF3** Ik kan al mijn werk voor school goed maken als ik maar genoeg tijd heb
- LV_SEF4** Ik kan bijna alles op school, als ik het maar blijf proberen
- LV_SEF5** Ik kan ook moeilijke dingen op school wel leren
- LV_SEF6** Ik weet zeker dat op school zelfs de moeilijkste taken me wel lukken

Uit de telling van het aantal valide antwoorden binnen leerlingen over de zes items, blijkt dat 14.112 leerlingen (%) alle items hebben beantwoord, dat 123 leerlingen (%) één item hebben overgeslagen en dat 146 leerlingen (%) geen enkel item valide hebben beantwoord. De overige 13 leerlingen (0.2%) hebben meerdere items overgeslagen. De betrouwbaarheid (α) bedraagt .85. Er zijn geen items die de betrouwbaarheid verlagen. Aan het bestand is een nieuwe variabele **LV_SEFFI** toegevoegd waarvan de score is berekend door het gemiddelde te nemen van de zes items, waarbij één itemscore mocht ontbreken. Op deze nieuwe variabele hebben 14.235 (%) leerlingen een score. Het gemiddelde bedraagt 3.55 ($SD = 0.66$).

9.3.7 Taakoriëntatie

Het concept ‘taakoriëntatie’ is door het Kohnstamm Instituut ingebracht en maakt ook deel uit van de vragenlijst die in COOL⁵⁻¹⁸-PO is afgenomen. Het komt inhoudelijk in vrij sterke mate overeen met de tweede orde schaal ‘mastery’ uit de ISM. In de theorie over ‘achievement goals’ wordt naast ‘taakoriëntatie’ het concept ‘ego-oriëntatie’ onderscheiden. Bij taakoriëntatie wordt succes gedefinieerd in termen van persoonlijke vooruitgang, bij ego-oriëntatie in termen van beter zijn dan anderen alsmede het graag tonen daarvan. De gebruikte itemset is afkomstig van Seegers, Van Putten en De Brabander (2002). Bij gebruik van deze schaal en/of items dient te worden verwezen naar deze publicatie.

In de vragenlijst waren de volgende vijf items als aparte vraag onder het tweede deel van de ISM-items geplaatst. Er was geen aparte inleiding. De antwoordmogelijkheden waren weer conform het eerdere ‘klopt’ regime. De volgende uitspraken werden voorgelegd:

LV_TSK1	Ik ben tevreden als ik op school iets heb geleerd dat ik begrijp
LV_TSK2	Ik maak liever moeilijke opdrachten waar ik iets nieuws van leer, dan gemakkelijke opdrachten
LV_TSK3	Ik vind het fijn wanneer ik op school iets heb geleerd dat ik belangrijk vind
LV_TSK4	Als ik op school iets niet meteen snap, ga ik er juist extra mijn best voor doen
LV_TSK5	Ik vind het fijn wanneer ik op school iets nieuws heb geleerd

Uit de telling van het aantal valide antwoorden binnen leerlingen over de vijf items blijkt dat 13.838 leerlingen (96.1%) alle items hebben beantwoord, dat 115 leerlingen (0.8%) één item hebben overgeslagen en dat 418 leerlingen (2.9%) geen enkel item valide hebben beantwoord. De overige 23 leerlingen (0.2%) hebben meerdere items overgeslagen. De betrouwbaarheid (α) van de schaal bedraagt .78. Er zijn geen items die de betrouwbaarheid verlagen. Aan het bestand is een nieuwe variabele **LV_TAAK** toegevoegd, waarvan de score is berekend door het gemiddelde te nemen van de vijf items, waarbij één itemscore mocht ontbreken. Op deze nieuwe variabele hebben 13.953 leerlingen (94.4%) een score. Het gemiddelde bedraagt 3.45 ($SD = 0.59$).

10 OUDERVragenLIJST

10.1 Inleiding

De voornaamste functie van de oudervragenlijst in het voortgezet onderwijs lag in het (opnieuw) verkrijgen van de achtergrondkenmerken van de leerlingen. In het onderzoeksvoorstel van COOL⁵⁻¹⁸ staat daarover opgemerkt: “Daarnaast krijgen de leerlingen een vragenlijst mee naar huis voor het verzamelen (bij de leerlingen die niet tot het cohort behoren) dan wel updaten van de gegevens over de achtergrond- en gezinskenmerken.” (p. 33 van het onderzoeksvoorstel). Met ‘het cohort’ wordt hier verwezen naar de meting van COOL⁵⁻¹⁸ in groep 8 van het basisonderwijs in 2011 en de meting in groep 5 van het basisonderwijs in 2008. Van de leerlingen die deelgenomen hebben aan de eerdere meting waren de achtergrondgegevens in principe al bekend.

Een tweede functie van de oudervragenlijst was het verkrijgen van aanvullende informatie van de ouders/verzorgers. In het onderzoeksvoorstel staat daarover: “Interessante terreinen zijn bijvoorbeeld het onderzoek naar opvoedstijlen en naar de mate waarin ouders een rolmodel zijn voor hun kinderen als het gaat om motivationele factoren en het aangaan van commitments.” (p. 6) Het onderzoeksvoorstel liet de nadere invulling van deze aanvullende informatie dus grotendeels in het midden – zie ook paragraaf 5.5 op p. 27 van het onderzoeksvoorstel, waar staat dat nadere beslissingen zullen worden genomen over ‘opvoedingsstijl’, ‘doeloriëntaties’ en ‘commitments’. Bij nadere beschouwing leek het wel zinvol om in de oudervragenlijst te vragen naar de opvoedstijl van de ouders/verzorgers, maar leek het te vergezocht om vragen over hun doeloriëntaties en commitments op te nemen. Het leek daarentegen wel zinvol om enkele vragen op te nemen die de ‘schooltijd’ van het kind betreffen.

Bijgevolg bestaat de oudervragenlijst uit twee delen, namelijk een deel met de formele achtergrondkenmerken, en een deel over o.a. opvoedstijl en schooltijd. Het eerste deel van de vragenlijst verenigt het beste van de oudervragenlijsten die eerder in het kader van PRIMA en VOCL werden afgenomen, en is in nauwe samenwerking tussen enkele van de onderzoekers tot stand gekomen. Het eerste deel van de vragenlijst is op enkele detailpunten na identiek aan de vragenlijst die drie jaar eerder bij de ouders van leerlingen in VO-3 is afgenomen, en aan de vragenlijst die in 2011 is afgenomen bij de ouders van leerlingen in groep 8 van het basisonderwijs.

Naast de genoemde concepten heeft het GION ook zelf een aantal concepten in het tweede deel van de vragenlijst opgenomen. De betreffende variabelen zijn niet in het databestand opgenomen en komen derhalve ook niet in dit rapport aan de orde. In paragraaf 10.2 schetsen we de gevolgde procedures. In paragraaf 10.3 bespreken we systematisch alle variabelen.

10.2 Procedures

De door de scholen naar Cito geretourneerde vragenlijsten zijn daar optisch ingelezen, volgens de wensen van het GION. Deze wensen betroffen het omgaan met dubbele antwoorden. Hoewel in de invulinstructie expliciet staat vermeld dat er per vraag slechts één antwoord mocht worden gegeven, is het de ervaring van het GION dat ook dan sommige respondenten bij sommige vragen toch meer dan één antwoord aankruisen (of omcirkelen). In veel van deze gevallen gaat het daarbij om twee aanliggende antwoorden op items die een Likert-achtig karakter hebben. De enig zinvolle interpretatie van dergelijke antwoorden is dat de respondent heeft willen aangeven dat zijn/haar antwoord “er tussen in” ligt. Aangezien dergelijke antwoordpatronen zinvol verwerkt kunnen worden, zijn ze ook met unieke codes ingevoerd. Dit wijkt af van de standaardbenadering bij optisch inlezen – en veelal ook bij handmatige invoer – waarbij dubbele antwoorden hetzij direct gelijk worden gesteld aan ‘geen antwoord’ of ‘fout antwoord’, hetzij één algemene code voor ‘dubbel antwoord’ krijgen toegekend. Na het inlezen zijn de data als een EXCEL-bestand aan het GION overgedragen. De verdere werkzaamheden zijn door het GION verricht. Deze werkzaamheden betreffen het omzetten naar SPSS, het toekennen van variabelenamen en het ‘labelen’ van de coderingen, het hercoderen van met name de coderingen die een dubbel antwoord representeren, het definiëren van ontbrekende data coderingen, het categoriseren van verbale antwoorden, die op sommige vragen gegeven konden worden, en het aanmaken van enkele schaalscores.

10.2.1 Invulinstructie

De vragenlijst begon met de volgende invulinstructie:

“In deze vragenlijst staan vragen over uw kind en over het gezin of huishouden waarin uw kind opgroeit.

Met ‘uw kind’ bedoelen we steeds *het kind dat in de tweede of derde klas VMBO, HAVO, of VWO onderwijs krijgt en deze vragenlijst mee naar huis heeft gebracht*. Als u meerdere kinderen in de tweede of derde klas heeft, krijgt u wellicht ook meer vragenlijsten. In dat geval vult u deze vragenlijst in voor het kind dat hierboven wordt genoemd.

De vragenlijst is bedoeld voor de ouders of verzorgers van het kind. Bij sommige vragen moet u een antwoord geven voor uzelf en voor uw partner. Met ‘uw partner’ bedoelen we de man of vrouw waarmee u *op dit moment* getrouwd bent of samenleeft. Het gaat dus steeds om het huishouden waarin u het kind opvoedt en verzorgt. Als er geen partner is – en u dus alleen voor uw kind zorgt – moet u alleen de vragen over uzelf beantwoorden. Vraag 7 is een uitzondering. Daar gaat het om de *biologische* grootouders van het kind.

Bij de meeste vragen hebben we de antwoorden al voorgegeven. Het enige dat u dan hoeft te doen, is een kruis in het hokje zetten dat het best past bij uw antwoord. Het liefst met een zwart potlood. Bij enkele antwoorden wordt nog om aanvullende informatie gevraagd; er staat dan ‘namelijk’ bij het antwoord. In die gevallen zet u een kruis in het hokje en schrijft u uw antwoord in vakje ernaast. Bij enkele vragen wordt u gevraagd een getal op te schrijven. Wilt u dan in elk hokje één cijfer zetten?

Wilt u overal maar *één antwoord* aankruisen? Als u meer dan 1 hokje aankruist, kan de computer dit antwoord niet lezen en is het antwoord voor ons verder onbruikbaar. Als u een potlood gebruikt kunt u vergissingen nog uitgummen.

In de vragen 15 en 20 worden twee keer 26 stellingen over verschillende eigenschappen van het kind gegeven. Met het aankruisen van de antwoorden geeft u als het ware een persoonlijke beschrijving van datgene wat uw kind uniek maakt. Probeer voor elke stelling een zo goed mogelijk antwoord te geven.”

De verdere verwerking van de gegevens die met de oudervragenlijst zijn verzameld, is *mutatis mutandis* gelijk aan de verwerking van de leerlingvragenlijst zoals die is beschreven in paragraaf 9.2. Met name voor het omgaan met ontbrekende waarden en dubbele coderingen verwijzen we naar hetgeen daar is vermeld. Hoewel in bovenstaande invulinstructie staat dat dubbele antwoorden niet verwerkt kunnen worden, is dat voor de VO-versie van de vragenlijst dus wel het geval. Voor de naamgeving van de variabelen hebben we bij de achtergrondkenmerken gekozen voor min of meer ‘intuïtieve’ namen, in plaats van een systematische naamgeving, waarbij de namen zouden beginnen met (bijvoorbeeld) OV, gevolgd door het vraagnummer.

10.2.2 Omgaan met onvolledige en/of inconsistente antwoorden

Hoewel de kwaliteit van de verkregen gegevens naar onze maatstaf op zijn minst als ‘behoorlijk goed’ kan worden gekwalificeerd, komen er zoals gebruikelijk ook onvolledige dan wel inconsistente antwoorden voor. Dat betreft combinaties van bij elkaar horende vragen. In de gevallen dat het antwoordpatroon op slechts één zinvolle wijze kon worden geïnterpreteerd, hebben we het op die wijze aangepast. Als voorbeeld nemen we vraag 3, die uit drie deelvragen bestond. De eerste daarvan luidde: “Zijn er in uw gezin nog andere thuiswonende kinderen dan het kind dat meedoet aan het onderzoek?”, met als antwoorden ‘nee’ en ‘ja’. In geval van het positieve antwoord werd doorverwezen naar de volgende twee vragen naar het aantal oudere en het aantal jongere kinderen. Indien de inleidende vraag niet was beantwoord, maar op minstens één van de twee vervolgvragen wel een aantal was aangegeven, is het ontbrekende antwoord op de inleidende vraag vervangen door ‘ja’. Op deze drie vragen komen echter ook echt inconsistente patronen voor, bijvoorbeeld antwoord ‘nee’ op de inleidende vraag, maar met toch een aangegeven aantal op minstens één van de vervolgvragen. Dergelijke inconsistenties zijn gehandhaafd.

10.2.3 Categoriseren open antwoorden

Op enkele vragen was er naast de ‘voorgegeven’ antwoorden voor de specifieke (a priori) antwoorden een extra antwoordmogelijkheid ‘anders, namelijk ...’. De respondenten die het betreffende antwoord aankruisten werden geacht om achter ‘namelijk’ een specificatie op te schrijven. Als voorbeeld kan de eerste vraag dienen: “Welke ouders of verzorgers zijn er in uw gezin?”, met als antwoordmogelijkheden: ‘moeder (of verzorgster) en vader (of verzorger)’, ‘alleen moeder (of verzorgster)’, ‘alleen vader (of verzorger)’, ‘twee ouders (of verzorgers) van hetzelfde geslacht’, ‘anders, namelijk ...’. Deze antwoorden hebben op de variabele **UDVER** de codes 1 tot en met 5. De verbale specificaties bij ‘anders, namelijk’ bleken hoofdzakelijk in drie categorieën te kunnen worden ondergebracht, namelijk ‘moeder met nieuwe partner’, ‘vader met nieuwe partner’ en ‘co-ouderschap’. De betreffende antwoorden zijn met codes 6, 7 en 8 opgenomen in de variabele **UDVER**. In deze gevallen is de oorspronkelijke code 5 dus vervangen door een van deze drie codes. Het bleek ook voor te komen dat een verbale specificatie eigenlijk op precies hetzelfde neerkwam als een van de vier voorgegeven antwoorden. Voor de betreffende respondenten is in dat geval code 5 vervangen door een van de eerste vier codes. De specificaties die niet goed in te delen waren zijn ondergebracht in de extra variabele **UDVER_S**. In de systematische beschrijving van deze en

soortgelijke variabelen richten we ons niet op het proces, maar op het uiteindelijke resultaat van de variabelen, zoals die in het bestand zijn opgenomen.

Het vragenlijstdeel van het SPSS bestand bevat de gegevens van 7989 ouders/verzorgers, die tenminste één vraag van de vragenlijst hebben beantwoord. In het totale SPSS bestand geeft de variabele **DUM_OV** aan of dit al dan niet het geval is. De in de rest van dit hoofdstuk vermelde percentages hebben betrekking op deze 7989 cases.

10.3 Systematische beschrijving

In paragraaf 10.3.1 geven we een systematische beschrijving van de achtergrondvariabelen uit het eerste deel van de vragenlijst. In paragraaf 10.3.2 documenteren we de ‘overige’ variabelen uit het tweede deel van de vragenlijst. Paragraaf 10.3.3 is gewijd aan de constructie van een variabele die de ‘sociaal-etnische’ achtergrond van de leerlingen aangeeft.

10.3.1 Achtergrondkenmerken

De eerste vijf vraagnummers vielen onder de noemer ‘Samenstelling gezin’. Er werd daarbij opgemerkt: *“Het gaat om de situatie zoals die op dit moment is.”*

1. Welke ouders of verzorgers zijn er in uw gezin?

- 1) moeder (of verzorgster) en vader (of verzorger)
- 2) alleen moeder (of verzorgster)
- 3) alleen vader (of verzorger)
- 4) twee ouders (verzorgers) van hetzelfde geslacht
- 5) anders, namelijk ...

In de vragenlijst stonden geen cijfers voor, maar antwoordhokjes achter de antwoorden. De cijfers verwijzen naar de codering van de antwoorden op deze variabele, die de naam **ODUVER** heeft. De vraag is niet beantwoord door 18 respondenten (0.2%). Uiteraard komt het eerste antwoord verreweg het meest voor (86.7%). De meeste van de verbale specificaties van ‘anders’ (code 5) konden hetzij worden ondergebracht in een van de voorgegeven categorieën, hetzij worden ingedeeld in een van de drie categorieën ‘moeder met nieuwe partner’, ‘vader met nieuwe partner’ en ‘co-ouderschap’. Deze categorieën hebben op **ODUVER** respectievelijk codes 6, 7 en 8 gekregen. NB: Inhoudelijk wijken de eerste twee van deze categorieën weinig af van de eerste voorgegeven categorie (vanwege de toevoeging “Het gaat om de situatie zoals die op dit moment is”). In totaal resteren 63 respondenten (0.8%) in de categorie ‘anders’. Van hen hebben 54 een specificatie gegeven. Deze specificaties zijn opgenomen in de variabele **ODUVER_S**. De overige 9 hebben geen specificatie opgeschreven (code 997 op **ODUVER_S**).

2. Wie vult de vragenlijst in?

- 1) moeder van het kind
- 2) vader van het kind
- 3) verzorgster van het kind
- 4) verzorger van het kind

De vraag is niet beantwoord door 30 respondenten (0,4%). Op grond van de gegeven dubbele antwoorden is op de variabele **INVULLER** aan deze vier categorieën een vijfde categorie toegevoegd: 'ouders samen'.

De volgende twee vragen betreffen het aantal andere kinderen in huis. Als eerste werd gevraagd:

3a. *Zijn er in uw gezin nog andere thuiswonende kinderen dan het kind dat meedoet aan het onderzoek?*

(1) nee

(2) ja

De naam van de variabele is **ANDKIND**. Degenen die 'ja' hadden geantwoord, werden doorverwezen naar vraag 3b, degenen die 'nee' hadden geantwoord naar vraag 4.

3b. *Hoeveel van die thuiswonende kinderen zijn ouder of even oud (tweeling) en hoeveel zijn jonger dan het kind dat meedoet aan het onderzoek?*

Deze vraag valt dus uiteen in twee sub-vragen. De variabelenamen zijn **NOUDK** voor het aantal oudere kinderen en **NJONGK** voor het aantal jongere kinderen. De antwoordmogelijkheden waren in beide gevallen: 0, 1, 2, 3, 4 en 5 of meer.

Wanneer op twee verschillende wijzen wordt gevraagd een situatie te beschrijven blijkt vaak dat de gegeven antwoorden elkaar niet volledig "dekken" of soms ronduit in tegenspraak zijn. Dat is ook hier het geval. Geprobeerd is om de inconsistenties in de antwoorden op vragen 3a, 3b1 en 3b2 in kaart te brengen en de antwoorden op elkaar af te stemmen. Enkele eenvoudige aanpassingen hebben te maken met ontbrekende antwoorden:

- Als op vraag 3a 'nee' was geantwoord, is aan het ontbrekende antwoord op vragen 3b1 en 3b2 code 8 toegekend, die gelabeld is als 'niet van toepassing' en als ontbrekend is gedefinieerd. Dit zijn dus 'terecht' ontbrekende antwoorden op deze twee vragen.
- Als het antwoord op vraag 3a ontbrak en op zowel vraag 3b1 als op vraag 3b2 0 was aangekruist, dan is het antwoord op vraag 3a op 'nee' gezet.
- Als het antwoord op vraag 3a ontbrak en op vragen 3b1 en 3b2 minstens eenmaal een aantal (anders dan 0) was aangekruist, dan is het antwoord op vraag 3a op 'ja' gezet.
- Als hetzij op vraag 3b1, hetzij op vraag 3b2 een aantal was aangekruist, maar het antwoord op de andere vraag ontbrak, dan is dat ontbrekende antwoord vervangen door 0.

Na deze aanpassingen zijn er nog 33 respondenten (0,4%) met een ontbrekend antwoord op **ANDKIND** en 40 respondenten (0,5%) met een ontbrekend antwoord op **NOUDK** en **JONGK**.

De volgende vraag luidde:

4. *Woont het kind ook nog een deel van de tijd in een ander gezin, bijvoorbeeld in verband met co-ouderschap na scheiding van de ouders?*

(1) nee

(2) ja

De naam van deze variabele is **ANDGEZ**. Op deze variabele komen 50 ontbrekende waarden voor (0.60.7%).

De volgende vraag luidde:

5. *Is het kind dat meedoet aan het onderzoek een adoptiekind of pleegkind?*

Apart voor adoptiekind en pleegkind kon ‘nee’ of ‘ja’ worden aangekruist. De beide antwoorden zijn gecombineerd tot één variabele **SRTKIND** met drie categorieën: ‘eigen kind’ (1), ‘adoptiekind’ (2), ‘pleegkind’ (3). Op deze variabele komen 82 ontbrekende waarden voor (1.0%).

De overige vragen van het eerste deel van de vragenlijst betroffen de achtergrondgegevens van de leerling. Erboven stond: *‘Achtergrondgegevens van de ouders of verzorgers’*. Daaronder stond als toelichting: “Bij de volgende vragen wordt vaak iets gevraagd over uzelf en uw partner. Met ‘uzelf’ bedoelen we degene die de vragenlijst invult; met ‘uw partner’ bedoelen we de man of vrouw met wie u op dit moment getrouwd bent of samenwoont. Als er geen partner is, hoeft u de vragen daarover niet in te vullen”.

Als eerste werd gevraagd:

6. *Waar zijn u en uw partner geboren? En waar het kind?*

- (1) Nederland
- (2) Suriname
- (3) Antillen/Aruba
- (4) Molukken
- (5) Turkije
- (6) Marokko
- (7) voormalig Joegoslavië
- (8) de voormalige Sovjet-Unie
- (9) Polen
- (10) overig Oost-Europa
- (11) China
- (12) Irak
- (13) Afghanistan
- (14) Somalië

Daarna volgden twee ‘open’ antwoordmogelijkheden. Eén vakje waar eventueel de antwoord op de vraag ‘een ander westers land, namelijk ...’ ingevuld kon worden en een tweede vakje waar het eventuele antwoord op ‘een ander niet-westers land, namelijk ...’ ingevuld kon worden. Gevraagd werd dit aan te kruisen (dan wel in te vullen) voor de invuller (uzelf), de partner en het kind.

Op de specificaties van de laatste twee antwoordmogelijkheden hebben we hetzelfde soort voorbewerkingen uitgevoerd als bij de eerste vraag over de samenstelling van het gezin. In een beperkt aantal gevallen was het mogelijk de specificatie onder te brengen in één van de voorgegeven antwoordcategorieën. Tevens hebben we evident foute antwoorden hersteld. Het kwam bijvoorbeeld voor dat ‘een ander westers land’ was ingevuld in het vakje voor niet-westers landen. Het omgekeerde geval kwam ook regelmatig voor.

Voor de grens tussen ‘westers’ en ‘niet-westers’ hebben we gebruik gemaakt van de definitielijsten die op de website van CBS zijn gepubliceerd. Australië en Nieuw-Zeeland zijn bij ‘westers’ ingedeeld. De landen Estland en Letland zijn niet bij; overig Oost-Europa, maar eveneens bij ‘westers’ ingedeeld. Zuid-Afrika is, in tegenstelling tot de afname in 2011, bij ‘niet-westers’ ingedeeld. Dat geldt ook voor personen waarvoor was ingevuld dat zij geboren waren in Indonesië, (voormalig) Nederlands Indië, of een van de Indonesische eilanden (Java, Borneo, etc.). We zijn ons bewust van het feit dat ouders en grootouders als westerse expats in deze gebieden aanwezig zijn geweest tijdens hun geboorte.

Behalve Nederland komen alleen Turkije, ‘anders, Westers’ en ‘anders, niet-Westers’ vaker dan 1% voor, en dan alleen bij de ouders/verzorgers. Bij de kinderen zelf komt alleen Nederland vaker dan 1% voor (namelijk 91.8%), gevolgd door ‘anders, Westers’ en ‘anders, niet-Westers’ (beide 0.7%). Gebruikers wordt aangeraden de geboortelanden naar eigen inzicht samen te voegen tot categorieën die afgestemd zijn op het beantwoorden van hun vraagstelling. Tevens zal het vaak geïndiceerd zijn om de (maximaal) drie geboortelanden te combineren tot één variabele, die de etnisch-culturele achtergrond van het kind betreft. In paragraaf 10.3.3 beschrijven we de constructie van een variabele die de sociaal-etnische achtergrond van de leerling aangeeft, waarbij gebruik is gemaakt van de geboortelanden. Het etnische aspect in deze variabele betreft echter alleen het onderscheid Westers versus niet-Westers.

De aangekruiste antwoorden zijn ondergebracht in drie variabelen **GBL_Z**, **GBL_P** en **GBL_K**, gelabeld als ‘geboorteland invuller’, ‘geboorteland partner’, en ‘geboorteland kind’. Het geboorteland van de invuller ontbreekt voor 26 respondenten (codering 99), terwijl 1 respondent een ‘overig’ antwoord gaf. Dat wil zeggen: het is onleesbaar of niet te interpreteren (codering 97) antwoord of er zijn dubbele antwoorden aangekruist (codering 98). Functioneel zijn er dus 27 ontbrekende waarden op **GBL_Z** (0.3%). Het geboorteland van de partner ontbreekt van 600 respondenten (codering 99) en het antwoord van 3 respondenten was niet te interpreteren. Functioneel zijn er dus 603 ontbrekende waarden op **GBL_P** (7.5%). Het geboorteland van het kind ontbreekt voor 420 respondenten en het antwoord van 3 respondenten was niet te interpreteren. Functioneel zijn er op **GBL_K** dus 423 ontbrekende waarden (5.3%).

De specificaties van de andere westerse en andere niet-westerse landen zijn ondergebracht in zes variabelen **GBL_ZOW**, **GBL_ZNW**, **GBL_POW**, **GBL_PNW**, **GBL_KOW**, en **GBL_KNW**, waarbij **OW** staat voor ‘overig westers’ en **NW** voor ‘overig niet-westers’. Deze variabelen hebben drie typen ontbrekende waarden. Codering 999 is gebruikt voor de respondenten die op de hoofdvraag code 97, 98 of 99 hebben gekregen, en is daarom gelabeld als ‘vraag niet beantwoord’. Codering 998 is gebruikt voor de respondenten die terecht niets hadden opgeschreven, en is daarom gelabeld als ‘niet van toepassing’. Codering 997 is gebruikt voor de respondenten die wel een specificatie hadden moeten opschrijven, maar dat of niet hebben gedaan of een niet leesbaar/interpreteerbaar antwoord invulden, en is daarom gelabeld als ‘geen specificatie’. Op de variabelen **GBL_ZOW**, **GBL_POW** en **GBL_KOW** zijn voor de overige westerse landen dezelfde coderingen gebruikt. Zo staat code 1 op alle drie variabelen voor Australië. Ook op de variabelen **GBL_ZNW**, **GBL_PNW** en **GBL_KNW** zijn voor de overige niet-westerse landen dezelfde coderingen gebruikt. Op deze variabelen staat code 1 voor Afrika (een nogal brede definitie), maar deze code komt voor het kind (**GBL_KNW**) niet voor. Vanaf code 2 worden specifieke landen benoemd.

Voor de invuller (**GBL_Z**, **GBL_ZOW**, **GBL_ZNW**) geldt:

- 27 maal (0.3%) is niets (bruikbaar) aangekruist of ingevuld.
- 117 maal (1.5%) is een ‘anders westers’ land ingevuld. Er zijn geen ontbrekende specificaties.
- 161 maal (2%) is ‘anders niet-westers’ ingevuld.

Voor de partner (**GBL_P**, **GBL_POW**, **GBL_PNW**) geldt:

- 603 maal (7.5 %) is niets (bruikbaar) aangekruist of ingevuld.
- 117 maal (1.5%) is ‘anders westers’ ingevuld.
- 202 maal (2.5%) is ‘anders niet-westers’ ingevuld

Voor het kind (**GBL_K**, **GBL_KOW**, **GBL_KNW**) geldt:

- 423 maal (4.2%) is niets (bruikbaar) aangekruist of ingevuld.
- 56 maal (0.7%) is ‘anders westers’ ingevuld.
- 58 maal (0.7%) is ‘anders niet-westers’ ingevuld.

De volgende vraag betrof de grootouders van het kind.

7. *Waar zijn de biologische grootouders van het kind geboren?*

Gevraagd werd dit aan te kruisen voor elk van de vier grootouders: moeder van moederszijde, vader van moederszijde, moeder van vaderszijde en vader van vaderszijde. De antwoordmogelijkheden waren hetzelfde als bij de vorige vraag, evenals de gevolgde procedure. Er zijn dus voor elk van de vier grootouders twee extra variabelen gemaakt waarin het resultaat van de bewerkingen die op de specificaties van ‘ander westers land’ en ‘ander niet-westers land zijn verricht, is opgeslagen. Deze acht variabelen hebben dezelfde drie typen ontbrekende waarden als de vergelijkbare variabelen uit de voorgaande vraag. Ook de coderingen die gebruikt zijn voor de overige westerse en overige niet-westerse landen zijn hetzelfde als de coderingen op de overeenkomstige variabelen voor de invuller, zijn/haar partner en het kind. De vier hoofdvariabelen hebben als namen: **GBL_MM**, **GBL_MV**, **GBL_VM** en **GBL_VV**. De acht toegevoegde variabelen hebben als namen: **GBL_MMOW**, **GBL_MMNW**, **GBL_MVOW**, **GBL_MVNW**, **GBL_VMOW**, **GBL_VMNW**, **GBL_VVOW**, en **GBL_VVNW**. Deze 12 variabelen kunnen desgewenst worden gecombineerd tot één variabele die de biologisch-etnische achtergrond van het kind aangeeft.

Voor moeders moeder (**GBL_MM**, **GBL_MMOW**, **GBL_MMNW**) geldt:

- Er zijn 47 (geen antwoord) + 1 (onduidelijk antwoord) = 48 functioneel ontbrekende antwoorden (0.6%).
- Er is 144 maal (1.8%) ‘anders westers’ ingevuld.
- Er is 287 maal (3.6%) ‘anders niet-westers’ ingevuld.

Voor moeders vader (**GBL_MV**, **GBL_MVOW**, **GBL_MVNW**) geldt:

- Er zijn 69 functioneel ontbrekende antwoorden (0.9%). Er waren daarbij geen onduidelijke antwoorden.
- Er is 130 maal (1.6%) ‘anders westers’ ingevuld.
- Er is 283 maal (3.5%) ‘anders niet-westers’ ingevuld.

Voor vaders moeder (**GBL_VM**, **GBL_VMOW**, **GBL_VMNW**) geldt:

- Er zijn 102 (geen antwoord) + 6 (onduidelijk antwoord) = 108 functioneel ontbrekende antwoorden (1.4%).
- Er is 133 maal (1.7%) ‘anders westers’ ingevuld. Bij 1 daarvan kon het land niet worden gespecificeerd door een onduidelijk antwoord.
- Er is 280 maal (3.5%) ‘anders niet-westers’ ingevuld. Bij 4 daarvan kon het land niet worden gespecificeerd door dubbele of onduidelijke antwoorden.

Voor vaders vader (**GBL_VV**, **GBL_VVOW**, **GBL_VVNW**) geldt:

- Er zijn 120 (geen antwoord) + 2 (onduidelijk antwoord) = 122 functioneel ontbrekende antwoorden (1.5%).
- Er is 129 maal (1.6%) ‘anders westers’ ingevuld. Bij 1 daarvan kon het land niet worden gespecificeerd door een onduidelijk antwoord.
- Er is 263 maal (3.3%) ‘anders niet-westers’ ingevuld. Bij 3 daarvan kon het land niet worden gespecificeerd door dubbele of onduidelijke antwoorden.

De volgende vraag betrof het aantal jaren dat de gezinsleden in Nederland wonen.

8a. *Hoeveel jaar wonen u en uw partner in Nederland?*

8b. *Hoeveel jaar woont uw kind in Nederland?*

Vraag 8a levert twee variabelen: **JARNED_Z** en **JARNED_P**, met als antwoordcategorieën: 1 = minder dan 3 jaar, 2 = 3 tot 5 jaar, 3 = 6 tot 9 jaar, 4 = 10 of meer jaar, 5 = altijd al. Vraag 8b levert de variabele **JARNED_K**, met iets andere antwoordmogelijkheden, namelijk: 1 = minder dan 2 jaar, 2 = 2 tot 5 jaar, 3 = 6 tot 8 jaar, 4 = meer dan 8 jaar, 5 = altijd al. Op deze drie variabelen komen respectievelijk 63 (0.7%), 890 (11.1%) en 50 (0.6%) functioneel ontbrekende antwoorden voor.

Vervolgens kwam de vraag of de ouders/verzorgers ooit asiel hadden aangevraagd.

9. *Hebben u of uw partner nu of in het verleden asiel in Nederland aangevraagd?*

(1) nee (2) ja

De variabelen heten **ASIEL_Z** en **ASIEL_P**. Op de eerste variabele komen 145 ontbrekende antwoorden (1.8%) voor, op de tweede 660 (8.3%).

De volgende vragen betreffen de gevolgde opleiding en de behaalde diploma's van invuller en partner. De vraag naar de gevolgde opleiding luidde als volgt:

10a. *Hieronder staan een aantal opleidingsniveaus. Wilt u het **hoogste** niveau aankruisen dat u en uw partner hebben **gevolgd**?*

- (1) geen onderwijs gevolgd
- (2) 1-3 jaar lager onderwijs/basisonderwijs
- (3) 4-6 jaar lager onderwijs/basisonderwijs
- (4) 1-2 jaar lager beroepsonderwijs (IBO/LBO/VBO)
- (5) 3-4 jaar lager beroepsonderwijs (IBO/LBO/VBO)
- (6) 1-2 jaar MULO/MAVO
- (7) 3-4 jaar MULO/MAVO
- (8) 1-3 jaar HAVO/HBS/MMS/VWO/atheneum/gymnasium
- (9) 4-6 jaar HAVO/HBS/MMS/ VWO/atheneum/gymnasium
- (10) middelbaar beroepsonderwijs (MBO/KMBO) of leerlingwezen
- (11) hoger beroepsonderwijs (HBO)
- (12) wetenschappelijk onderwijs (universiteit)

Deze vraag werd als volgt toegelicht: *“Als u niet het precieze type weet (bijvoorbeeld omdat het om een avondopleiding gaat of omdat de opleiding in het buitenland is gevolgd), probeer dan een zo goed mogelijke inschatting te geven. Kleuteronderwijs niet meetellen.”* Gevraagd werd dus om zowel voor de invuller als voor de partner het hoogste gevolgde opleidingsniveau aan te kruisen. Het blijkt echter dat een groot aantal respondenten zich hier niet aan heeft gehouden, en meer dan één antwoord heeft aangekruist. De meervoudige antwoorden zijn naar de hoogst voorkomende waarde gehercodeerd. Dus bijvoorbeeld 267 is gehercodeerd naar 7. Op deze variabelen – **HOOPLG_Z** en **HOOPLG_P** – komen bijgevolg geen meervoudige antwoorden voor. Bij de constructie van de sociaal-etnische variabele (paragraaf 10.3.3) is gebruik gemaakt van deze twee variabelen. Voor de invuller zelf ontbreekt het antwoord 48 maal (0.6%), voor de partner zijn er 697 ontbrekende antwoorden (8.7%).

De vraag naar de behaalde diploma's luidde als volgt:

10b. *Van welk onderwijstype hebben u en uw partner het diploma behaald?*

- (1) geen enkel diploma
- (2) lager beroepsonderwijs (IBO/LBO/VBO)
- (3) MULO/MAVO
- (4) HAVO/HBS/MMS/VWO/atheneum/gymnasium
- (5) middelbaar beroepsonderwijs (MBO/KMBO) of leerlingwezen
- (6) hoger beroepsonderwijs (HBO)
- (7) wetenschappelijk onderwijs

Eronder werd opgemerkt dat er meerdere antwoorden konden worden gekozen. Voor de invuller is de vraag niet beantwoord door 59 respondenten (0.7%), bij de partner zijn er 751 ontbrekende antwoorden (9.4%). Hoewel het dus toegestaan was meerdere antwoorden aan te kruisen, zijn ook in dit geval in het bestand twee variabelen – **HODIPL_Z** en **HODIPL_P** – opgenomen met daarin uitsluitend het hoogst behaalde diploma van invuller en partner.

De volgende vraag betrof het hebben van een betaalde baan:

11. *Hebben u en uw partner een betaalde baan voor 12 uur of meer per week?*

Voor zowel de invuller als de partner kon 'ja' of 'nee' worden aangekruist. Voor de invuller is de vraag niet beantwoord door 82 respondenten (1.0%), voor de partner zijn er 733 ontbrekende antwoorden (9.2%). De twee variabelen heten **BETWER_Z** en **BETWER_P**.

De volgende vraag betrof de religieuze gezindheid van de gezinsleden en luidde als volgt:

12. *Tot welke kerk of geloof rekenen u en uw partner zich? En tot welke rekt u het kind?*

- (1) tot geen enkele kerk of geloof
- (2) Rooms-Katholieke kerk
- (3) Protestantse Kerk in Nederland (inclusief Ned. Hervormde Kerk; Gereformeerde Kerken in Nederland; Evangelisch Luthers)
- (4) Protestants-Orthodoxe kerk (o.a. Gereformeerde Kerken Vrijgemaakt; Christlijk Gereformeerde Kerken; Gereformeerde Gemeenten)
- (5) Pinkster- en Evangeliegemeenten
- (6) Islam
- (7) een andere Christelijke kerk, namelijk ...
- (8) een andere kerk of geloof, namelijk ...

Bij 'een andere Christelijke kerk' en 'een andere kerk of geloof' diende, indien van toepassing, achter 'namelijk' de naam van die kerk of dat geloof te worden ingevuld. De antwoorden die hier zijn gegeven zijn op dezelfde wijze bewerkt als de open antwoorden bij de vragen naar de geboortelanden. Vervolgens zijn voor elk van de drie gezinsleden twee nieuwe variabelen gemaakt waarin de specificaties van 'een andere Christelijke kerk' en 'een andere kerk of geloof' zijn opgeslagen. De hoofdvariabelen hebben als namen **GLF_Z**, **GLF_P** en **GLF_K**. De variabelen met de specificaties hebben als namen **GLF_ZAC**, **GLF_ZAG**, **GLF_PAC**, **GLF_PAG**, **GLF_KAC**, en **GLF_KAG**. De laatste twee letters – AC en AG – staan voor Ander Christelijk en Ander Geloof

Voor de invuller geldt:

- Er zijn 80 (geen antwoord) + 21 (dubbel antwoord) = 101 functioneel ontbrekende antwoorden (1.3%)
- Er is 77 maal (1.0%) 'andere Christelijke kerk' aangekruist
- Er is 85 maal (1.1%) 'andere kerk of geloof' aangekruist

Voor de partner geldt:

- Er zijn 740 (geen antwoord) + 9 (dubbel antwoord) = 749 functioneel ontbrekende antwoorden (9.4%)
- Er is 67 maal (0.8%) 'andere Christelijke kerk' aangekruist
- Er is 73 maal (0.9%) 'een andere kerk of geloof' aangekruist

Voor het kind geldt:

- Er zijn 472 (geen antwoord) + 23 (dubbel antwoord) = 495 functioneel ontbrekende antwoorden (6.2%)
- Er is 73 maal (0.9%) 'andere Christelijke kerk' aangekruist
- Er is 68 maal (0.9%) 'andere kerk of geloof' aangekruist

De laatste vragen van het deel met achtergrondkenmerken betroffen de taal of talen die in het gezin worden gesproken en de ervaren beheersing van de Nederlandse taal. De eerste van deze vragen betrof de taal die het kind het meest spreekt:

13a. *Welke taal spreekt het kind het meeste?*

- (1) Nederlands
- (2) Fries/streektaal/dialect
- (3) buitenlandse taal
- (4) niet van toepassing

Daaronder werden vier 'gesprekspartners' genoemd, namelijk 'uzelf', 'uw partner', 'broers of zussen' en 'vriendjes of vriendinnetjes'. Achter iedere gesprekspartner kon de meest gebruikte taal worden aangekruist. Op deze vier variabelen, met de namen **TAAL_KZ**, **TAAL_KP**, **TAAL_KBZ**, en **TAAL_KV**, komen relatief veel dubbele antwoorden voor. Het gaat daarbij vooral om de combinaties van Nederlands met Fries/streektaal/dialect en van Nederlands met een buitenlandse taal. Deze twee combinaties hebben op de vier variabelen codes 5 en 6 gekregen. De (weinige) andere dubbele combinaties hebben code 8 'overig' gekregen, die als ontbrekend is gedefinieerd. Voor 'uzelf' is de vraag niet beantwoord door 28 respondenten en zijn er 10 overige antwoorden. Voor de partner zijn er 596 ontbrekende en 9 overige antwoorden, voor broers of zussen 376 ontbrekende en 10 overige antwoorden en voor vriendjes of vriendinnetjes 243 ontbrekende en 6 overige antwoorden. Functioneel ontbreekt daarmee het antwoord in respectievelijk 0,5%, 7.6%, 4,8% en 3.1% van de gevallen. De aantallen 'niet van toepassing'-antwoorden zijn respectievelijk 48, 120, 224, 54. De categorie 'niet van toepassing' was opgenomen voor het geval dat er bijvoorbeeld geen partner zou zijn. De meeste ontbrekende oordelen voor de partner hadden eigenlijk als 'niet van toepassing' moeten voorkomen. Het kan uiteraard ook voorkomen dat een kind geen broers of zussen heeft of geen vriend(inn)en. Er zijn echter ook 48 'niet van toepassing' oordelen op **TAAL_KZ**, hetgeen zou impliceren dat de invuller van de vragenlijst niet met het kind communiceert.

De volgende vraag luidde:

13b. *Welke taal spreken u en uw partner het meeste met elkaar?*

Er waren dezelfde antwoordmogelijkheden als op vraag 13a. Op deze variabele **TAAL_ZP** komen 264 (geen antwoord) + 12 (dubbel antwoord) = 276 (3.5%) functioneel ontbrekende antwoorden voor en is 301 maal (3.9%) 'niet van toepassing' geantwoord.

De laatste vraag van dit deel van de vragenlijst betrof de beheersing van de Nederlandse taal door de respondent en de partner op vier verschillende aspecten, namelijk verstaan/begrijpen, spreken, lezen, schrijven. De vraag luidde:

14. *In welke mate beheersen u en uw partner de Nederlandse taal?*

Apart voor 'uzelf' en voor de partner stonden de vier aspecten onder elkaar. De antwoordmogelijkheden waren als volgt: 1) niet/zeer slecht, 2) slecht, 3) redelijk, 4) goed, 5) zeer goed. De variabelenamen zijn respectievelijk **NED_ZV**, **NED_ZP**, **NED_ZL** en **NED_ZS** voor de respondent, en **NED_PV**, **NED_PP**, **NED_PL** en **NED_PS**. [De laatste letters V, P, L en S staan voor verstaan, praten, lezen en schrijven.] Op deze variabelen komen in beperkte mate dubbele antwoorden voor. In verreweg de meeste gevallen betrof dit een 'aanliggend dubbel' antwoord. Deze gevallen zijn gehercodeerd naar 1.5, 2.5, 3.5 en 4.5. In een paar gevallen betrof het een anderszins dubbel antwoord. Deze gevallen zijn gecodeerd als 8, welke code als ontbrekend is gedefinieerd.

Er zijn 38 respondenten die geen van de vier op hem/haar zelf betrekking hebbende vragen hebben beantwoord. Daarnaast zijn er respectievelijk 6, 49, 48, 59 specifiek ontbrekende antwoorden. Er zijn 691 respondenten die geen van de vier op de partner betrekking hebbende vragen hebben beantwoord. Daarnaast zijn er voor de partner respectievelijk 3, 41, 40, 45 specifiek ontbrekende antwoorden.

De correlaties tussen deze items zijn erg hoog. De gemiddelde correlatie tussen de items die op de respondent zelf betrekking hebben bedraagt .86, hetgeen tot een betrouwbaarheid van .96 leidt. Voor de vier items die betrekking hebben op de partner resulteren eveneens .85 en .95.

Er zijn twee schaaltes **NEDZELF** en **NEDPART** gemaakt met een samenvattende score van de oordelen over het beheersingsniveau van het Nederlands. Vanwege de zeer hoge correlaties tussen de items is een schaalscore gedefinieerd indien minstens twee van de vier oordelen beschikbaar waren. Op **NEDZELF** resulteert een score voor 7904 respondenten (98,9%), met een gemiddelde van 4.69 en een standaard deviatie van 0.53. Op **NEDPART** resulteert een score voor 7258 partners (90,8%), met een gemiddelde van 4.62 en een standaard deviatie van 0.62.

We menen te mogen stellen dat het deel van de vragenlijst over de achtergrondkenmerken in het algemeen 'goed' is ingevuld. Behalve bij de vragen die op een partner betrekking hebben, is het percentage ontbrekende antwoorden (vrij) laag. Verreweg de meeste ontbrekende antwoorden op de vragen over de partner zijn terug te voeren op de afwezigheid van een partner. Er zijn ook relatief weinig dubbele antwoorden waarvoor geen zinvolle hercodering mogelijk was.

10.3.2 Overige variabelen

Het eerste concept is de opvoedstijl. Het betreft de visie van de respondent op in hoeverre het kind zelf beslist over een aantal onderwerpen, waarover wel eens een heftige woordenwisseling tussen ouders/verzorgers en opgroeiende jongeren ontstaan. De vraag luidde:

16. *Wie neemt de beslissingen over de volgende onderwerpen?*

Daaronder stonden de volgende items:

- OV_AUT1** Hoe laat het kind thuis moet zijn
- OV_AUT2** Of het kind een (bij)baantje mag hebben
- OV_AUT3** Of het kind thuis alcohol mag drinken
- OV_AUT4** Of het kind op feestjes alcohol mag drinken
- OV_AUT5** Welke tv-programma's het kind kijkt
- OV_AUT6** Welke vervolgopleiding het kind gaat volgen
- OV_AUT7** Wanneer het kind zijn/haar huiswerk maakt
- OV_AUT8** Hoe lang het kind internet en/of computerspellen speelt
- OV_AUT9** Of het kind uit mag gaan

Bij elk onderwerp waren de antwoordmogelijkheden: 1 = ik en/of partner, **zonder** overleg met het kind, 2 = ik en/of partner, **na** overleg met het kind, 3 = ik en/of partner **samen** met het kind, 4 = het kind zelf, **na** overleg met mij en/of mijn partner, 5 = het kind zelf, **zonder** overleg met mij en/of mijn partner, 6 = niet van toepassing. De items zijn gebaseerd op de literatuur over 'opvoedstijlen' en betreffen het concept 'authoritatieve' opvoedstijl. Wikipedia geeft hiervan de volgende omschrijving: "Authoritative parenting is characterized by a child-centered approach that holds high expectations of maturity, compliance to parental rules and directions, while allowing for an open dialogue about those rules and behaviors between parent and child". Geschikte verwijzingen naar dit concept zijn Baumrind (1978) en Steinberg, Elmen en Mounts (1989). De betreffende items zijn door Veenstra ingebracht in VOCL'93 (zie Veenstra & Kuyper, 2004). Daar betrof het vragen vanuit het perspectief van de leerlingen.

Uit een telling van het aantal bruikbare antwoorden (in de range 1 tot en met 5) blijkt dat 4528 respondenten (56,7%) alle items bruikbaar hebben beantwoord, dat 995 (12,5%) één onbruikbaar antwoord hebben gegeven, 1089(13,6%) twee onbruikbare antwoorden, etc. Er zijn 155 respondenten (1,9%) die geen enkel item bruikbaar hebben beantwoord

De betrouwbaarheid van de itemset, die hier berekend is in de groep van 4528 respondenten die op alle items een antwoord in de range 1 tot en met 5 hebben gegeven, is .76. Er zijn geen betrouwbaarheidsverlagende items. Als er maximaal één ontbrekende itemscore wordt toegestaan, zouden slechts 5523 ouders (69,1%) een schaalscore krijgen. Daarom is hier een soepeler regel gehanteerd: naast maximaal één echt ontbrekende (of onduidelijke) itemscore zijn maximaal vier "niet van toepassing" antwoorden toegestaan. Aan dit soepeler criterium voldoen 7417 respondenten (92,8%). Aan het bestand zijn twee nieuwe variabelen toegevoegd. **OV_AUTON** geeft de mate van de volgens de respondent toegestane autonomie van het kind aan, dat wil zeggen de mate waarin de respondent heeft aangegeven een 'laissez faire' opvoedstijl te hanteren. De score op deze schaal is berekend door het gemiddelde van de (minimaal vier) itemscores te bepalen. Het gemiddelde op deze schaal bedraagt 3,05 ($SD = 0,60$). De tweede variabele (**OV_AUTH**) kan geacht worden de 'authoritatieve' opvoedstijl te representeren. De score berust op een telling van het aantal antwoorden 'het kind zelf, na overleg met mij of mijn partner',

welk antwoord als kenmerkend voor deze opvoedstijl wordt gezien. Vanwege de verschillen in het aantal bruikbare itemscores is echter niet de frequentie genomen, maar de (relatieve) proportie. Als bijvoorbeeld 1 item niet was beantwoord, er geen 'niet van toepassing' antwoorden waren gegeven, en het betreffende antwoord was driemaal gegeven, dan is de schaalscore $3/8 = 0.38$. Als er 1 ontbrekend en 1 'niet van toepassing' antwoord zou zijn, dan zou de schaalscore $3/7 = 0.43$ zijn. Er is hetzelfde criterium gehanteerd als bij de berekening van **OV_AUTON**: er mag op maximaal vijf items een niet valide score voorkomen en daarvan mag hoogstens één echt ontbreken (dan wel 'onduidelijk' zijn). Het gemiddelde op deze variabele bedraagt 0,17 ($SD = 0,18$).

De overige vragen betroffen de lestijd op school, de tijd die het kind aan huiswerk besteedde en spijbelen. De eerste vraag (**LESTIJD**) luidde:

Hoeveel lesuren per week heeft het kind? Alle vakken samen, inclusief praktijklessen.

- (1) minder dan 20 uur
- (2) 20-21 uur
- (3) 22-23 uur
- (4) 24-25 uur
- (5) 26-27 uur
- (6) 28-29 uur
- (7) 30 uur of meer
- (8) weet niet

De vraag is niet of onduidelijk beantwoord door 444 respondenten (5,6%). De antwoordmogelijkheid 'weet niet' is door 527 respondenten (6,6%) aangekruist.

De volgende vraag betrof de hoeveelheid tijd die buiten school werd besteed aan taken die met school samenhangen (huiswerk). Deze vraag luidde: "*Hoeveel uren per week werkt het kind buiten schooltijd aan de volgende schooltaken?*" Daaronder stonden de volgende drie items:

HW_MAKEN	Het maken van huiswerk
HW_LEREN	Het leren van proefwerken of schriftelijke overhoringen
HW_OPDRA	Het maken van opdrachten en werkstukken

De antwoordmogelijkheden waren: 1 = niet van toepassing, 2 = 0 uur, 3 = 1-2 uur, 4 = 3-4 uur, 5 = 5-6 uur, 6 = 7-8 uur, 7 = 9-10 uur, 8 = 11-12 uur, 9 = 13-14 uur, 10 = 15 uur of meer, 11 = weet niet. In totaal hebben 131 respondenten (1,6%) voor geen enkele van deze taken iets ingevuld, met andere woorden zij hebben de vraag in het geheel niet beantwoord. Hiervoor is code 99 gebruikt. Daarnaast komen per taak specifieke ontbrekende waarden voor, respectievelijk 5 (<0.1%), 11 (0.1%) en 65 (0.8%). Hiervoor is codering 98 gebruikt. De meest voor de hand liggende interpretatie van de specifieke ontbrekende waarden is dat ze hetzelfde betekenen als het antwoord 'niet van toepassing' dan wel '0 uur'. Gebruikers van het bestand kunnen desgewenst zelf een dergelijke hercodering aanbrengen. Codering 97 is gebruikt voor onduidelijke, dat wil zeggen niet-aanliggend dubbele antwoorden, en komt respectievelijk 41, 11 en 13 maal voor.

De laatste twee vragen betroffen het spijbelen door het kind. Als eerste (**SPB_FREQ**) werd gevraagd naar de frequentie.

Hoe vaak heeft het kind volgens u dit schooljaar gespijbeld?

- (1) nooit

- (2) 1 of een enkele keer
- (3) elke maand wel een keer
- (4) elke week wel een keer
- (5) elke week wel 2 of 3 keer
- (6) zowat elke dag
- (7) weet niet

Deze vraag is niet beantwoord door 62 respondenten (0.8%); 10 respondenten hebben een onduidelijk, niet-aanliggend dubbel antwoord gegeven, en 126 respondenten hebben ‘weet niet’ geantwoord. De overige 7791 respondenten (97.5%) hebben een frequentie oordeel gegeven. In bijna 86% was het antwoord ‘nooit’.

Ten slotte werd gevraagd naar de duur van het spijbelen (**SPB_DUUR**):

Als het kind spijbelt, hoe lang is dat meestal?

- (1) een lesuur
- (2) een halve dag
- (3) een hele dag
- (4) meerdere dagen
- (5) weet niet

De vraag is niet beantwoord door 6566 respondenten (82.2%); 12 respondenten hebben een onduidelijk antwoord gegeven, en 197 hebben ‘weet niet’ geantwoord. De ontbrekende antwoorden (6566) zijn gegeven door respondenten die de vorige vraag met ‘nooit’ hadden beantwoord. Het ontbrekende antwoord betekent in dat geval ‘niet van toepassing’. Verder zijn er 164 respondenten bij die de eerdere vraag nooit hadden beantwoord, maar bij deze vraag wel een spijbelduur hebben ingevuld. Een kruistabel van de antwoorden op beide vragen laat zien dat het invullen niet helemaal consistent is gedaan.

10.3.3 Constructie SOCETN (sociaal-etnische achtergrond)

De variabele die de ‘sociaal-etnische achtergrond’ van de leerling aangeeft, is door het ITS en KI ontwikkeld ten behoeve van PRIMA. De variabele is gebaseerd op twee vragen in de oudervragenlijst. De eerste vraag betrof de geboortelanden van de respondent en diens partner (indien aanwezig) – **GBL_Z** en **GBL_P**. De tweede vraag betrof het opleidingsniveau van deze perso(o)n(en) – **HOOPLG_Z** en **HOOPLG_P**. Als er een verschil was tussen de geboortelanden van de respondent en de partner, gaf het geboorteland van de moeder (of eventueel de verzorgster) de doorslag. Daartoe is gebruik gemaakt van de vraag wie de vragenlijst invulde (**INVULLER**). Bij de opleidingen ging het om de hoogst gevolgde opleiding door een van beide personen, ongeacht of dat de vader (verzorger) dan wel de moeder (verzorgster) was. Tabel 10.1 toont de oorspronkelijke antwoorden op de vraag naar de geboortelanden (variabelen **GBL_Z** en **GBL_P**).

De opgeschreven specificaties van ‘een ander westers land, namelijk ...’ en ‘een ander niet-westers land, namelijk ...’ zijn geïnventariseerd en gesystematiseerd. Sommige van deze specificaties konden direct worden ondergebracht in een van de andere voorgegeven antwoordmogelijkheden. De 13 bewerkte categorieën zijn vervolgens gedichotomiseerd tot ‘Westers’ versus ‘niet-Westers’, hetgeen overeenkomt met het onderscheid ‘autochtoon’ versus ‘allochtoon’ in de terminologie van ITS/KI (zie Driessen, Mulder,

Ledoux, Roeleveld, & Veen, 2009; bladzijde 31). De Molukken is hierbij tot Westers gerekend, maar Suriname en Antillen/Aruba tot niet-Westers.

Tabel 10.1

Geboortelanden respondent en partner (N = 7.989)

	respondent	partner
Nederland	7240	6617
Suriname	69	73
Antillen/Aruba	34	29
Molukken	1	2
Turkije	137	148
Marokko	58	58
voormalig Joegoslavië	23	27
voormalige Sovjet-Unie	15	15
Polen	17	13
overig Oost-Europa	10	10
China	37	32
Irak	14	16
Afghanistan	22	20
Somalië	7	7
ander Westers land	117	117
ander niet-Westers land	161	202
overig antwoord	1	3
geen antwoord	26	603

Bij de tweede vraag moest van 12 categorieën van opleidingen de hoogste worden aangekruist die enerzijds de respondent en anderzijds de partner had gevolgd. Tabel 10.2 toont de antwoorden op deze vraag.

Tabel 10.2

Hoogst gevolgde opleiding door respondent en partner (N = 7.989)

	respondent	partner
geen onderwijs	42	53
1-3 jaar lager onderwijs/basisonderwijs	22	27
4-6 jaar lager onderwijs/basisonderwijs	52	81
1-2 jaar lager beroepsonderwijs (LBO/VBO/VMBO-beroeps)	73	98
3-4 jaar lager beroepsonderwijs (LBO/VBO/VMBO-beroeps)	588	885
1-2 jaar MULO/MAVO/VMBO-theoretisch (+ gemengd)	93	102
3-4 jaar MULO/MAVO/VMBO-theoretisch (+ gemengd)	679	584
1-3 jaar HAVO/HBS/VWO/atheneum/gymnasium	130	111
4-6 jaar HAVO/HBS/VWO/atheneum/gymnasium	494	314
middelbaar beroepsonderwijs (MBO/KMBO) of leerlingwezen	2681	2198
hoger beroepsonderwijs (HBO)	2286	1938
wetenschappelijk onderwijs (universiteit)	801	901
geen antwoord	48	697

Net als in het bestand van de eerste dataverzameling van COOL⁵⁻¹⁸ in het VO hebben wij de posities ‘4-6 jaar HAVO/HBS/MMS/VWO/atheneum/gymnasium’ en ‘middelbaar beroepsonderwijs (MBO/KMBO) of leerlingwezen’ verwisseld. Voor een onderbouwing hiervan, gebaseerd op de gemiddelde intelligentie van de leerlingen in de onderscheiden opleidingscategorieën van de ouders, verwijzen we naar paragraaf 11.3.3 in het betreffende technische rapport (Zijsling et al. 2013). De resulterende categorieën zijn vervolgens getrichotomiseerd in ‘laag’, ‘midden’ en ‘hoog’. De categorie ‘laag’ bevat de eerste vijf categorieën van gevolgde opleiding en kan worden aangeduid als ‘hoogstens LBO gevolgd’. De categorie ‘midden’ bevat de volgende drie categorieën van gevolgde opleiding, aangevuld met middelbaar beroepsonderwijs, en kan worden aangeduid als ‘hoogstens MBO gevolgd’. De categorie ‘hoog’ bevat de bovenbouw van HAVO en VWO, HBO en WO. Zoals opgemerkt is vervolgens de hoogst voorkomende categorie van de twee ouders/verzorgers genomen.

De ‘sociaal etnische achtergrond’ van de leerlingen is vervolgens gebaseerd op de kruistabel van de driedeling in onderwijsniveaus met ‘Westers versus niet-Westers’. Deze kruistabel ziet er als volgt uit (Tabel 10.3).

Tabel 10.3

Sociaal-etnische achtergrond van de leerlingen

opleidingsniveau	niet-Westers	Westers	totaal
laag	130	368	498
midden	195	2616	2811
hoog	228	4408	4636
totaal	553	7392	7945

Er zijn 44 leerlingen (0.6%) die niet op de gecombineerde variabele (**SOCETN**) kunnen worden ingedeeld. In totaal zijn het middelste en – meer nog – het hoogste opleidingsniveau het meest gevuld. Bij de niet-Westerse leerlingen komen de drie niveaus echter ongeveer even vaak voor. De sociaal-etnische variabele is geordend van ‘niet-Westers-laag’ tot en met ‘Westers-hoog’.

11 BESCHRIJVING DATABESTAND

Het databestand bevat de toets- en vragenlijstgegevens van de 16297 leerlingen die aan ten minste één van de onderdelen van de gegevensverzameling in 2014 in VO-3 hebben deelgenomen. In de voorgaande hoofdstukken is al een groot deel van de variabelen van het databestand geïntroduceerd. Enkele administratieve variabelen zijn nog niet besproken, maar de precieze betekenis van deze variabelen kan eenvoudig worden afgeleid uit de ‘variable labels’ en ‘value labels’ in het databestand. Onderstaand overzicht geeft een korte toelichting bij elke variabele uit het databestand. Op basis van de informatie in de hoofdstukken 2 tot en met 10, de samenvattende informatie in dit hoofdstuk en beide soorten labels moet het naar onze inschatting mogelijk zijn om probleemloos met het databestand te werken. Bij gebruik van de variabelen uit het databestand dient gerefereerd te worden naar de bron die is genoemd in het hoofdstuk waar de variabele beschreven wordt.

Administratieve variabelen

LL_ID	Leerlingidentificatie
SCH_ID	Schoolidentificatie
KLAS_ID	Klasidentificatie
LEERJAAR	Leerjaar
TARGET	Heeft de leerling eerder aan COOL ⁵⁻¹⁸ deelgenomen?
LL_ID_0708	Nummer van de targetleerling in de PO5-afname van 2007/08
LL_ID_1011	Nummer van de targetleerling in de PO8-afname van 2010/11
DEELNAME	Heeft de school ‘collectief’ of ‘individueel’ deelgenomen?
DUM_INT	Is de intelligentietoets gemaakt?
DUM_BGL	Is de begrijpend lezen toets gemaakt?
DUM_TVZ	Is de toets taalverzorging gemaakt?
DUM_WIS	Is de wiskundetoets gemaakt?
DUM_ENG	Is de toets Engels gemaakt?
DUM_BSC	Is deelgenomen aan de vragenlijst burgerschapscompetenties?
DUM_LV	Is de leerlingvragenlijst beantwoord?
DUM_OV	Is de oudervragenlijst beantwoord?
NTOETS	Aan hoeveel onderdelen is deelgenomen?

Variabelen beschreven in hoofdstuk 2

DAT_INT	Op welke datum is de intelligentietoets gemaakt?
DAT_BGL	Op welke datum is de begrijpend lezen toets gemaakt?
DAT_TVZ	Op welke datum is de toets taalverzorging gemaakt?
DAT_WIS	Op welke datum is de wiskundetoets gemaakt?
DAT_ENG	Op welke datum is de toets Engels gemaakt?
DAT_LV	Op welke datum is de leerlingvragenlijst gemaakt?
DAT_OV	Op welke datum is de oudervragenlijst gemaakt?
MIS_BGL	Percentage ontbrekende antwoorden bij de begrijpend lezen toets
MIS_TVZ	Percentage ontbrekende antwoorden bij de toets taalverzorging
MIS_WIS	Percentage ontbrekende antwoorden bij de wiskundetoets
MIS_ENG	Percentage ontbrekende bij de toets Engels

SCH_REG	In welke provincie staat de school?
SCH_STED	Urbanisatiegraad van de school
SEKSE	Geslacht leerling
LEEFTIJD	Leeftijd van de leerling op 01-03-2014
OW_TYPE	Welk type onderwijs volgt de leerling?
OW_SECTOR	Voor welke sector heeft de leerling gekozen?
OW_PROFIEL	Voor welk profiel heeft de leerling gekozen?

Variabelen beschreven in hoofdstuk 3

DUM_IQSCO1	Heeft de leerling een intelligentiescore? (criterium 1)
DUM_IQSCO2	Heeft de leerling een intelligentiescore? (criterium 2)
INT_VAL1	Aantal antwoorden onderdeel 1
INT_VAL2	Aantal antwoorden onderdeel 2
INT_VAL3	Aantal antwoorden onderdeel 3
INT_VAL4	Aantal antwoorden onderdeel 4
INT_VAL5	Aantal antwoorden onderdeel 5
INT_SOM1	Somscore onderdeel 1
INT_SOM2	Somscore onderdeel 2
INT_SOM3	Somscore onderdeel 3
INT_SOM4	Somscore onderdeel 4
INT_SOM5	Somscore onderdeel 5
INT_SCO1	Intelligentiescore 1
INT_SCO2	Intelligentiescore 2

Variabelen beschreven in hoofdstuk 4

BGL_VERS	Welke toetsversie heeft de leerling gemaakt?
BGL_RSCO	Ruwe score begrijpend lezen
BGL_WSCO	Gewogen ruwe score begrijpend lezen
BGL_THET	OPLM vaardigheidsschatting begrijpend lezen
BGL_BANK3	OPLM bankscore begrijpend lezen

Variabelen beschreven in hoofdstuk 5

TVZ_VERS	Welke toetsversie heeft de leerling gemaakt?
TVZ_RSCO	Ruwe score taalverzorging
TVZ_BANK3	OPLM bankscore taalverzorging

Variabelen beschreven in hoofdstuk 6

WIS_VERS	Welke toetsversie heeft de leerling gemaakt?
WIS_RSCO	Ruwe score wiskunde
WIS_WSCO	Gewogen ruwe score wiskunde
WIS_THET	OPLM vaardigheidsschatting wiskunde
WIS_BANK3	OPLM bankscore wiskunde

Variabelen beschreven in hoofdstuk 7

ENG_VERS	Welke toetsversie heeft de leerling gemaakt?
ENG_RSCO	Ruwe score Engels
ENG_WSCO	Gewogen ruwe score Engels
ENG_THET	OPLM vaardigheidsschatting Engels

Variabelen beschreven in hoofdstuk 8

BS_ATOT	attitude totaal
BS_VTOT	vaardigheid totaal
BS_RTOT	reflectie totaal
BS_KTOT	kennis totaal
BS_ADH	attitude democratisch handelen totaal
BS_ADH_1	attitude democratisch handelen: ieders stem willen horen
BS_ADH_2	attitude democratisch handelen: kritische bijdrage willen leveren
BS_AMV	attitude maatschappelijke verantwoordelijkheid
BS_ACO	attitude omgaan met conflicten
BS_AVE	attitude omgaan met verschillen
BS_VDH	vaardigheid democratisch handelen totaal
BS_VDH_1	vaardigheid democratisch handelen: opkomen voor eigen mening
BS_VDH_2	vaardigheid democratisch handelen: luisteren naar mening van anderen
BS_VMVCO	vaardigheid maatschappelijke verantwoordelijkheid en omgaan met conflicten
BS_VVE	vaardigheid omgaan met verschillen
BS_RDH	reflectie democratisch handelen
BS_RMV	reflectie maatschappelijke verantwoordelijkheid
BS_RCO	reflectie omgaan met conflicten
BS_RVE	reflectie omgaan met verschillen
BS_KDH	kennis democratisch handelen
BS_KMV	kennis maatschappelijke verantwoordelijkheid
BS_KCO	kennis omgaan met conflicten
BS_KVE	kennis omgaan met verschillen
BS_SW	sociale wenselijkheid

Variabelen beschreven in hoofdstuk 9

LV_WDO1	De docenten weten meestal wel hoe ik me voel
LV_WDO2	Ik kan met de docenten over problemen praten
LV_WDO3	Als ik ongelukkig voel, kan ik daar met docenten praten
LV_WDO4	Ik voel mij bij de docenten op mijn gemak
LV_WDO5	De docenten begrijpen mij
LV_WDO6	Ik heb goed contact met de docenten
LV_WDO7	Ik zou liever andere docenten hebben
LV_WDOC	Contact met docenten
LV_LES1	Hoeveel lesuren per week moet je op school zijn?
LV_LES2A	#uren p.wk - krijg je les van een docent?
LV_LES2B	#uren p.wk - moet je opdrachten/werkstukken maken olv een docent?
LV_LES2C	#uren p.wk - moet je huiswerk/proefwerk maken olv een docent?
LV_LES2D	#uren p.wk - mag je (in klas) zelf weten wat je doet?
LV_LES2E	#uren p.wk - vallen uit (ziekte docent/andere reden)?
LV_BF_FACT1	Extraversie
LV_BF_FACT2	Mildheid
LV_BF_FACT3	Ordelijkheid
LV_BF_FACT4	Emotionele Stabiliteit
LV_BF_FACT5	Intellectuele Autonomie

LV_BF_NFACT	Aantal valide factorscores
LV_WKL1	Ik heb veel contact met klasgenoten
LV_WKL2	Ik zou liever in een andere klas zitten
LV_WKL3	Wij hebben een leuke klas
LV_WKL4	Ik kan goed met mijn klasgenoten overweg
LV_WKL5	In mijn klas voel ik me soms alleen
LV_WKL6	Ik vind het leuk om met leerlingen uit mijn klas om te gaan
LV_WKLAS	Omgang met klas
LV_ISM01	Ik werk op school het liefst zoveel mogelijk samen met anderen
LV_ISM02	Ik doe tijdens de les mijn best een beloning van de leraar/lerares te krijgen
LV_ISM03	Ik wil graag dat het nuttig is wat ik op school leer
LV_ISM04	Ik wil graag de beste zijn
LV_ISM05	Voor interessant schoolwerk doe ik beter mijn best
LV_ISM06	Als je op school 'bonuspunten' zou krijgen, zou ik beter mijn best doen
LV_ISM07	Ik vind het belangrijk een beloning te krijgen voor goed schoolwerk
LV_ISM08	Ik help graag andere leerlingen goede cijfers te halen
LV_ISM09	Het is voor mij belangrijk van vriend/innen complimenten te krijgen voor mijn schoolwerk
LV_ISM10	Ik ben alleen maar tevreden als ik een van de besten van de klas ben
LV_ISM11	Ik voel mij betrokken bij andere leerlingen op school
LV_ISM12	Ik werk op school het beste als ik complimenten krijg
LV_ISM13	Ik probeer vaak de leider van een groepje te zijn
LV_ISM14	Ik zie graag dat mijn schoolwerk steeds beter wordt
LV_ISM15	Ik help graag anderen met schoolwerk, ook als ik zelf niet zo goed ben
LV_ISM16	Ik krijg graag de kans iets opnieuw te doen om het te verbeteren
LV_ISM17	Ik doe mijn best op school zodat ik de leider van een groepje word
LV_ISM18	Als mijn schoolwerk vooruit gaat, doe ik nog beter mijn best
LV_ISM19	Ik doe mijn best op school als ik met anderen samen werk
LV_ISM20	Ik ben op school graag de leider van een groepje
LV_ISM21	Ik krijg graag een compliment voor goed schoolwerk
LV_ISM22	Ik werk harder als ik beter dan anderen probeer te zijn
LV_ISM23	Ik doe goed mijn best op school omdat ik mijn schoolwerk interessant vind
LV_ISM24	Het is voor mij belangrijk om van mijn ouders een compliment te krijgen
LV_ISM25	Ik baal ervan als mijn vriend/innen op school slechte cijfers halen
LV_ISM26	Bij moeilijke problemen doe ik juist beter mijn best
LV_ISM27	Ik vind het belangrijk dat leerlingen elkaar helpen op school
LV_ISM28	Ik werk op school liever samen met anderen dan alleen
LV_ISM29	Ik doe goed mijn best om nieuwe dingen op school te begrijpen
LV_ISM30	Ik probeer altijd mijn schoolwerk nog beter te doen
LV_ISM31	Het is voor mij belangrijk om van leraar/es compliment te krijgen voor mijn schoolwerk
LV_ISM32	Ik wil op school beter presteren dan mijn klasgenoten
LV_ISM33	Als ik op school beloningen zou krijgen zou ik beter mijn best doen
LV_TSK	ISM - Task
LV_EFF	ISM - Effort
LV_COMP	ISM - Competition
LV_SOCO	ISM - Social concern
LV_SOPO	ISM - Social power
LV_PRS	ISM - Praise
LV_AFF	ISM - Affiliation

LV_TOK	ISM - Token
LV_MAST	ISM - Mastery
LV_PERF	ISM - Performance
LV_SOCL	ISM - Social
LV_EXTR	ISM - Extrinsic
LV_GMOT	ISM - Motivatie overall (strikt)
LV_GMOT2	ISM - Motivatie overall (soepel)
LV_SEF1	Ik weet zeker dat dit jaar alles op school wel zal lukken
LV_SEF2	Ik kan op school zelfs de moeilijkste opdrachten maken als ik mijn best doe
LV_SEF3	Ik kan al mijn werk voor school goed maken als ik maar genoeg tijd heb
LV_SEF4	Ik kan bijna alles op school, als ik het maar blijf proberen
LV_SEF5	Ik kan ook moeilijke dingen op school wel leren
LV_SEF6	Ik weet zeker dat op school ook de moeilijkste taken me wel lukken
LV_SEFFI	Self-efficacy
LV_TSK1	Ik ben tevreden als ik iets heb geleerd dat ik begrijp
LV_TSK2	Ik maak liever moeilijke opdrachten waar ik iets nieuws van leer dan gemakkelijke
LV_TSK3	Ik vind het fijn als ik iets heb geleerd wat ik belangrijk vind
LV_TSK4	Als ik iets niet meteen snap, ga ik er juist extra mijn best voor doen
LV_TSK5	Ik vind het fijn als ik op school iets nieuws heb geleerd
LV_TAAK	Taakorientatie

Variabelen beschreven in hoofdstuk 10

OUDVER	Welke ouders/verzorgers in uw gezin
OUDVER_S	Ouders in gezin, anders...
INVULLER	Wie vult vragenlijst in?
ANDKIND	Nog andere kinderen thuis?
NOUDK	#oudere kinderen
NJONGK	#jongere kinderen
ANDGEZ	Woont kind ook in ander gezin?
SRTKIND	Type kind
GBL_Z	Gbl. invuller
GBL_ZOW	Gbl. invuller ov. westers
GBL_ZNW	Gbl. invuller ov. niet-westers
GBL_P	Gbl. partner
GBL_POW	Gbl. partner ov. westers
GBL_PNW	Gbl. partner ov. niet-westers
GBL_K	Gbl. kind
GBL_KOW	Gbl. kind ov. westers
GBL_KNW	Gbl. kind ov. niet-westers
GBL_MM	Gbl. moeders moeder
GBL_MMOW	Gbl. moeders moeder ov. westers
GBL_MMNW	Gbl. moeders moeder ov. niet-westers
GBL_MV	Gbl. moeders vader
GBL_MVOW	Gbl. moeders vader ov. westers
GBL_MVNW	Gbl. moeders vader ov. niet-westers
GBL_VM	Gbl. vaders moeder
GBL_VMOW	Gbl. vaders moeder ov. westers
GBL_VMNW	Gbl. vaders moeder ov. niet-westers

GBL_VV	Gbl. vaders vader
GBL_VVOW	Gbl. vaders vader ov. westers
GBL_VVNW	Gbl. vaders vader ov. niet-westers
JARNED_Z	Hoe lang in NL (invuller)?
JARNED_P	Hoe lang in NL (partner)?
JARNED_K	Hoe lang in NL (kind)?
ASIEL_Z	Ooit asiel aangevraagd (invuller)?
ASIEL_P	Ooit asiel aangevraagd (partner)?
HOOPLG_Z	Hoogste opl.niveau invuller
HOOPLG_P	Hoogste opl.niveau partner
HODIPL_Z	Hoogste diploma invuller
HODIPL_P	Hoogste diploma partner
BETWER_Z	Betaalde baan >12 uur (invuller)?
BETWER_P	Betaalde baan >12 uur (partner)?
GLF_Z	geloof/kerk invuller
GLF_ZAC	Andere chr. kerk (invuller)
GLF_ZAG	Andere kerk/geloof (invuller)
GLF_P	geloof/kerk partner
GLF_PAC	Andere chr. kerk (partner)
GLF_PAG	Andere kerk/geloof (partner)
GLF_K	geloof/kerk kind
GLF_KAC	Andere chr. kerk (kind)
GLF_KAG	Andere kerk/geloof (kind)
TAAL_KZ	taal kind - invuller
TAAL_KP	taal kind - partner
TAAL_KBZ	taal kind - broers/zussen
TAAL_KV	taal kind - vriend(inn)en
TAAL_ZP	taal invuller - partner
NED_ZV	invuller: NI verstaan/begrijpen
NED_ZP	invuller: NI spreken
NED_ZL	invuller: NI lezen
NED_ZS	invuller: NI schrijven
NEDZELF	Beheersing NI zelf
NED_PV	partner: NI verstaan/begrijpen
NED_PP	partner: NI spreken
NED_PL	partner: NI lezen
NED_PS	partner: NI schrijven
NEDPART	Beheersing NI partner
OV_AUT1	Hoe laat het kind thuis moet zijn
OV_AUT2	Of kind een bijbaantje mag hebben
OV_AUT3	Of kind thuis alcohol mag drinken
OV_AUT4	Of kind op feestjes alcohol mag drinken
OV_AUT5	Naar welke tv-programma's kind kijkt
OV_AUT6	Welke vervolgopleiding kind gaat volgen
OV_AUT7	Wanneer kind huiswerk maakt
OV_AUT8	Hoe lang internet en/of computerspellen speelt
OV_AUT9	Of kind uit mag gaan
OV_AUTON	Autonoom beslissen

OV_AUTH	'Authoritative' opv.stijl
LESTIJD	Hoeveel lesuren per week heeft het kind?
HW_MAKEN	#uren p.wk - maken van huiswerk
HW_LEREN	#uren p.wk - leren van proefwerken/schriftelijke overhoring
HW_OPDRA	#uren p.wk - maken van opdrachten en werkstukken
SPB_FREQ	Hoe vaak dit schooljaar gespijbeld?
SPB_DUUR	Als het kind spijbelt, hoe lang dan?
SOCETN	Sociaal-etnische achtergrond

LITERATUUR

- Batenburg, Th. A. van, & Werf, M.P.C. van der (2004). *NSCCT: Niet Schoolse Cognitieve Capaciteiten Test. Voor groep 4, 6 en 8 in het basisonderwijs. Verantwoording, normering en handleiding*. Groningen: GION.
- Baumrind, D. (1978). Parental disciplinary patterns and social competence in children. *Youth and Society*, 9, 238-276.
- Blom G. (1958). *Statistical estimates and transformed beta-variables*. Wiley: New York.
- Boertien, H., Heuves, T., & Kremers, E.J.J. (1995). *Constructie toetsen Nederlands en wiskunde voor VOCL-cohortonderzoek 1993*. Arnhem: Cito.
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296-322.
- Cito (2007a). *Toets 3 vmbo-bb volg- en adviessysteem: handleiding voor de toetsafname op papier*. Arnhem: Cito.
- Cito (2007b). *Toets 3 vmbo-kb/gt volg- en adviessysteem: handleiding voor de toetsafname op papier*. Arnhem: Cito.
- Cito (2007c). *Toets 3 havo/vwo volg- en adviessysteem: handleiding voor de toetsafname op papier*. Arnhem: Cito.
- Driessen, G., Mulder, L., & Roeleveld, J. (2012). *Cohortonderzoek COOL5-18. Technisch rapport basisonderwijs, tweede meting 2010/11*. Nijmegen: ITS, Radboud Universiteit Nijmegen ; Amsterdam: Kohnstamm Instituut.
- Driessen, G. Mulder, L., Ledoux, G., Roeleveld, J. & Veen, H. van der (2009). *Cohortonderzoek COOL5-18. Basisrapport basisonderwijs, eerste meting 2007/08*. Nijmegen/Amsterdam: ITS/SCO-KI.
- Eggen, T.J.H.M. (1993). Itemresponstheorie en onvolledige gegevens. In T.J.H.M Eggen & P.F. Sanders (Eds.), *Psychometrie in de Praktijk*. Arnhem: Cito.
- Eggen, T.J.H.M. (2004). *Contributions to the theory and practice of computerized adaptive testing*. Arnhem: Cito.
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP.
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (Commissie Meijerink) (2008). *Over de drempels met taal en rekenen*. Enschede: SLO.
- Feldt, L., & Brennan, R. (1989). Reliability. In: R. Linn (Ed.), *Educational measurement* (3rd ed.). The American Council on Education: MacMillan.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hendriks, A. A. J., Hofstee, W. K. B., & De Raad, B. (1999). *Handleiding bij de Five-Factor Personality Inventory (FFPI)*. Lisse: Swets Test Publishers.

- Hendriks, A. A. J., Hofstee, W. K. B., & De Raad, B. (2011). Handleiding bij de Five Factor Personality Inventory II (FFPI-II). Houten: Bohn Stafleu van Loghum
- Hendriks, A. A. J., Kuyper, H., Offringa, G. J., & Werf, M. P. C. van der (2008). Assessing Young Adolescents' Personality With the Five-Factor Personality Inventory. *Assessment*, 15, 304-316.
- Janssen, J., Schoot, F. van der, & Hemker, B. (2005). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 4*. Arnhem: Cito.
- Keuning, J. (2004). *De ontwikkeling van een beoordelingssysteem voor het beoordelen van computer based tests*. Arnhem: Cito.
- Keuning, J., Zijssling, D., Naaijer, H., & Timmermans, A. (2015). *Cohortonderzoek COOL5-18: Technisch rapport meting havo-5 in 2013*. Groningen: GION onderzoek/onderwijs.
- Kolen, M.J. & Brennan, R.L. (1995). *Test equating: Methods and practices*. New York: Springer-Verlag.
- Kuyper, H., & Werf, M.P.C. van der (2005). VOCL'99-3: *Prestaties en opvattingen van leerlingen in de derde klas van het voortgezet onderwijs*. Groningen: GION.
- Lee, W. & Ban, J. (2010). A comparison of IRT linking procedures. *Applied Measurement in Education*, 23, 23-48.
- Lindgren, B.W. (1976) *Statistical theory* (3rd ed.). New York: Macmillan.
- Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McInerney, D.M. & Ali, J. (2006). Multidimensional and hierarchical assessment of school motivation: cross-cultural validation. *Educational Psychology*, 26 (6), 717-734
- McInerney, D. M. & Sinclair, K. E. (1991). Cross-cultural model testing inventory of school motivation. *Educational and psychological measurement*, 51, 123-133.
- Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor, MI: University of Michigan.
- Keizer-Mittelhaeuser, M-A., Naaijer, H., Zijssling, D., & Timmermans, A. (2015). *Cohortonderzoek COOL5-18: Technisch rapport meting vwo-6 in 2014*. Groningen: GION onderzoek/onderwijs.
- Muraki, E., Hombo, C. M. , & Lee, Y.-W. (2000). Equating and linking of performance assessments. *Applied Psychological Measurement*, 24(4), 325-337.
- Peetsma, T. T. D., Wagenaar, E., & Kat, E. de (2001). School motivation, future time perspective and well-being of high school students in segregated and intergated schools in the Netherlands and the role of ethnic self-description. In J. Koppen, I. Lunt, & C. Wulf (Eds.). *Education in Europe; Culture, Values, Institutions in transition* (pp. 54-74). Münster/New York: Waxmann.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295.
- Steinberg, L., Elmen, J. D., & Mounts, N. S. (1989). Authoritative parenting, psychosocial maturity, and academic success among adolescents. *Child Development*, 60, 1424-1436.
- Ten Dam, G., Geijssel, F., Reumerman, R., & Ledoux, G. (2011). Measuring Young People's Citizenship Competences. *European Journal of Education*, 46 (3), 354-372.
- Thissen, D., & Wainer, H. (2001). *Test scoring*. Philadelphia: Lawrence Erlbaum Associates.

- Van Til, A. (2011). *Verantwoording meting taal en rekenen*. Arnhem: Cito.
- Veenstra, D., & Kuyper, H. (2004). Effective Students and Families: The Importance of Individual Characteristics for Achievement in High School. *Educational Research and Evaluation*, 10, 41-70.
- Veldhuijzen, N. H., Goldebelt, P., Sanders, P. M. (1993): *Klassieke testtheorie en generaliseerbaarheidstheorie*. 1 ed. Arnhem: CITO, Instituut voor Toetsontwikkeling.
- Verhelst, N.D., (1993). Itemresponstheorie. In: T.J.H.M. Eggen & P.F. Sanders (red.). *Psychometrie in de praktijk*. (pp. 83-178). Arnhem: Cito.
- Verhelst, N. D., Glas, C. A. W. (1995). The generalized one parameter model: OPLM. In: G.H.Fischer&I.W.Molenaar (eds.). *Rasch models: their foundations, recent developments and applications*. (pp.215-238). New York: Springer.
- Verhelst N.D., Glas C.A.W., &Verstralen H.H.F.M. (1995). *OPLM: One-Parameter Logistic Model. Computer program and manual*. Arnhem: Cito.
- Webb, N., Shavelson, R., &Haertel, E. (2007). Reliability coefficient and generalizability theory. In: C. Rao & S. Sinharay (Eds.), *Handbooks of Statistics 26: Psychometrics*. The Netherlands: Elsevier.
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational evaluation*, 31, 114-128.
- Zijsling, D., Kuyper, H., Lubbers, M. J., & Werf, M. P. C. van der (2005). *VOCL '99-3. Technisch Rapport*. Groningen: GION.
- Zijsling, D., Keuning, J., Kuyper, H., Batenburg, T. van, &Hemker, B. (2009). *Technisch rapport eerste meting van COOL⁵⁻¹⁸ in het derde leerjaar van het voortgezet onderwijs*. Groningen: GION.
- Zijsling, D., Keuning, J., Kuyper, H., Batenburg, T. van, & Hemker, B. (2013). *Technisch rapport meting VO-3 in 2008 (tweede, herziene versie)*. Groningen: GION.

BIJLAGEN

1 Dienstrooster Eetcafé De Lommer

Dienstrooster Eetcafé De Lommer				
	Murat	Lianne	Hans	Ingrid
Maandag	B-dienst	A-dienst	Vrij	C-dienst
Dinsdag	Vrij	C-dienst	B-dienst	A-dienst
Woensdag	Vrij	B-dienst	C-dienst	A-dienst
Donderdag	B-dienst	Vrij	A-dienst	C-dienst
Vrijdag	A-dienst	B-dienst	C-dienst	Vrij
Zaterdag	A-dienst	Vrij	C-dienst	Vrij
Zondag	B-dienst	C-dienst	Vrij	A-dienst
Toelichting diensten: A-dienst 07.30 - 16.00 B-dienst 09.30 - 18.00 C-dienst 14.30 - 22.00				

2 De slimme tabaksfabrikant

De slimme tabaksfabrikant

Trucjes van tabaksfabrikanten

(1) Wist jij dat er suiker in een sigaret zit? In sommige sigaretten zit zelfs honing of chocolade. Veel mensen weten dat niet. Hoe kunnen ze dat ook weten? Op een pot jam staat meer informatie over de inhoud dan op een pakje sigaretten!

(2) Tabaksbladeren hebben een bittere, vieze smaak. Eigenlijk zou niemand die willen roken. Daarom worden sigaretten lekkerder gemaakt. Hoe? Door smaakjes aan de tabak toe te voegen. Maar liefst 5 tot 20% van een sigaret bestaat uit toegevoegde stoffen! Er wordt bijvoorbeeld suiker, honing, zoethout, chocolade, vanille en mint toegevoegd. Maar ook vruchtensappen, zoals limoen, sinaasappel en ananas.

(3) Roze pakjes voor de meiden en zwarte pakjes voor de jongens. Zelfs de sigaretten in het pakje hebben een kleurtje. Allemaal trucjes om jongeren de sigaretten te laten kopen.

Je gelooft het misschien niet, maar fabrikanten geven sigaretten zelfs een kleurtje om ze voor jongeren aantrekkelijk te maken. Zo zijn er bijvoorbeeld roze en zwarte sigaretten. Ze zien er heel onschuldig uit. Het lijken haast wel 'speelgoed sigaretjes'. Door hun opvallende kleur zijn deze pakjes populair onder rokende jongeren. Maar deze sigaretten zijn niet zo onschuldig als ze eruit zien. Ze zijn net zo schadelijk als elke andere sigaret.

(4) Gekleurde pakjes en gekleurde sigaretten zijn trucjes van fabrikanten om jongeren te verleiden om de sigaretten te kopen. Uit onderzoek blijkt dat veel mensen denken dat lichtgekleurde pakjes minder ongezond zijn dan de donker gekleurde pakjes. Ook denken veel mensen dat een pakje met een gouden afbeelding minder ongezond is dan een pakje met een rode afbeelding. Maar alle sigaretten zijn ongezond. Het uiterlijk van een pakje sigaretten kan mensen dus op het verkeerde been zetten.

(5) Vroeger kon de tabaksindustrie zich alles veroorloven. Sigarettenmerken maakten op vreemde manieren reclame. Niks was te gek. Er bestaat een advertentie waarin een baby zijn moeder vraagt om wat meer te roken. Daar wordt ze namelijk rustiger van. Bijna niemand had het lef om de fabrikanten tegen te spreken. Je kunt je wel voorstellen dat er dus veel mensen rookten.

(6) Eind jaren '50 rookten alle Amerikaanse mannen van 15 jaar en ouder. Met dank aan het sterkste wapen ter wereld, dat zonder beperkingen mocht worden gebruikt: reclame. Wie rookte was volwassen, stoer en hoorde erbij. Je kent vast wel de Marlboro-man: een stoere cowboy, genietend van zijn vrijheid. Met een peuk in z'n mond zat hij op z'n paard.

De tabaksindustrie wist in de jaren '50 al dat je van roken ziek wordt. De bewijzen stapelden zich op. Maar de grote sigarettenmerken trokken zich hier niets van aan. Ze zeiden zelfs dat je kon roken 'zonder gevaar'. 'Niks aan de hand', leken ze te zeggen. 'Steek er gerust nog eentje op'. Nu kan de tabaksindustrie een dergelijke aansporing heus niet meer kwijt.

(7) Sigarettenmerken proberen van alles om reclame te maken. Acteurs krijgen grote sommen geld als ze een bepaald merk sigaret roken in hun films. Nu er meer regels zijn over reclame maken voor tabak, wordt het lastig voor de tabaksindustrie en probeert ze op andere manieren de aandacht te trekken.

(8) De tabaksindustrie ging over op sponsoring. Bijvoorbeeld van sportevenementen en muziekfestivals. Merken gaven er geld voor om op deze evenementen en festivals hun naam te laten zien. Maar ook dit mag nu niet meer. Daarom hebben ze weer wat nieuws bedacht: brand stretching. Wat dat is? Het betekent zoiets als het oprekken van je merk.

(9) Oftewel: Marlboro ging kleding op de markt zetten. Camel kwam met stoere schoenen. Heel handig, want hiervoor waren nog geen regels. Zo kreeg je zeilwedstrijden met als sponsor Pall Mall Pleasure Wear. En met Camel schoenen was je opeens een ontdekkingsreiziger.

(10) Onze regering springt bepaald niet voorzichtig om met de belangen van de tabaksindustrie. Op elk pakje rookwaar staat verplicht een zwartomrande waarschuwing in grote zwarte letters zoals: "Roken tijdens de zwangerschap is slecht voor uw baby", "Roken is dodelijk" en "Roken verouderd de huid". Die waarschuwingen doen denken aan een bekend spreekwoord: "De een zijn dood, is de ander zijn brood". Voor de tabaksindustrie moet je dat dan wel omdraaien: "De een zijn brood, is de ander zijn dood". Dat die "dood" een lekker smaakje heeft, mooi verpakt is of stiekem geadverteerd wordt, maakt hem echter niet mooier.

[http://www.stivoro.nl/Voor_jongeren/Werkstukinfo/Trucjes_van_tabaksfabrikanten]

13-01-2010

ISBN 978-90-367-9444-2 (eBook)
NUR 840
© 2017 GION Onderwijs / Onderzoek



Grote Rozenstraat 3
9712 TG Groningen



zeker weten